BIOCHEMICAL CHARACTERIZATION OF FIVE GH130-FAMILY ENZYMES FROM *Caldanaerobius polysaccharolyticus* ATCC BAA-17 AND INSIGHTS ON THEIR METABOLIC ROLE AND REACTION MECHANISMS

BY

HANS MÜLLER PAUL

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Animal Sciences
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Master's Adviser:

Professor Isaac Cann

ABSTRACT

Proteins in the glycoside hydrolase family 130 (GH130, CAZy database) have been proposed to perform the phosphorolysis of β-1,2 and β-1,4-mannosyl linkages between the mannose at the non-reducing end of substrates and mannose, glucose or *N*-acetylglucosamine residues, with the subsequent release of α-mannose-1-phosphate. In this study, we compare five different GH130 enzymes (CpMan130 A-E) encoded within the genome of *Caldanaerobius polysaccharolyticus*, a thermophilic anaerobic bacterium able to ferment mannan as the sole carbon source.

Analysis of substrate specificity and end product release allowed for the identification of pathways involving GH130 enzymes in the metabolism of mannans with different structures by this organism. Mechanistic studies involving the binding order and amino acid mapping on a three-dimensional model of Man130B helped to elucidate the ordered sequential bi-bi mechanism utilized by these enzymes. Phylogenetic analysis of over 950 sequences assigned to the GH130 family, combined with differences in amino acid conservation and substrate specificity, revealed a new subgroup for this family, GH130_3, consisting of thermostable enzymes that act on β-1,2-linked manno-oligosaccharides. The genomic context of all genes in the proposed subgroup GH130_3 suggests that they appear in pairs, preceded by an ABC-like transporter.

ACKNOWLEDGEMENTS

This project would not have been possible without the support of many people. I am very thankful to my adviser, Dr. Isaac Cann, for the unique opportunity to be part of his group and work in the state of the art facilities at the Energy Biosciences Institute, as well as for his guidance and for the many valuable lessons I have learned from him. I am also thankful to my committee members, Dr. Roderick Mackie and Dr. Jason Ridlon, for their time, patience and support. This research would not be possible without the infrastructure and financial support from the Energy Biosciences Institute and the Carl R. Woese Institute for Genomic Biology, for which I am very thankful, and the fellowships from CAPES[1] and the Science without Borders program, and the Lemann Foundation, without which I would not have had the opportunity to study abroad.

I would like to thank Dr. Leandro Tessler for his encouragement and support in this endeavor. I am grateful to my family, for their effort in giving me access to the excellent education that allowed me to reach this point, and for their support in enduring the many challenges I had to face through life and which shaped who I have become. I thank my friends, old and new, who have become an integral part of my life and with whom I have shared many moments, both good and bad, and all the students and post-docs in the Cann, Mackie and Ridlon labs, from whom I have learned a great deal, discussed results with and called upon for help so many times. I would also like to thank my girlfriend Júlia, for all her love and support during this long process, and for enduring

[1] Fellowship process number BEX 13769-13-8

the hardship of a long distance relationship for the larger part of the two years we have been separately working our way into a life together.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 Introduction

Plant cell wall polysaccharides are the most abundant organic compounds found in nature, and are an important energy source for a variety of organisms. Due to its complex and recalcitrant nature [1, 2], the utilization of the carbohydrate components of these polymers require a series of reactions to be debranched and depolymerized to monomeric units, which can then be shuttled into the central metabolism [3, 4]. Such reactions are carried out by carbohydrate active enzymes (CAZymes, grouped in the CAZy database) [3-6], which can be roughly divided into four main groups. Although polysaccharide lyases (PLs) and carbohydrate esterases (CEs) are sufficiently divergent from glycoside hydrolases (GHs) to be classified in their own CAZyme families, the same does not hold true for glycoside phosphorylases (GPs), which can be found within both the glycosyltransferase (GTs) and glycoside hydrolase families [7].

The proteins in the glycoside hydrolase family 130 (GH130) have been described based on the identification of 4-O-β-D-mannosyl-D-glucose phosphorylase activity (EC 2.4.1.281) for a protein derived from the gene BF0772 of *Bacteroides fragilis* (BfMGP) [8]. Up to now, proteins in this family have been proposed to perform the phosphorolysis of β-1,2 and β-1,4-mannosyl linkages between mannose and mannose, glucose or *N*-acetylglucosamine residues at the non-reducing end of substrates [7, 9-12]. A unique proton relay mechanism for GH130 enzymes has been proposed on the basis of the three-dimensional structure of BfMGP, suggesting that α-mannose-1-phosphate is afforded by a single displacement mechanism leading to inversion of anomeric

configuration [9, 13]. Some GH130 proteins which do not display the three conserved basic residues that comprise the phosphate binding site have been proposed to function as glycoside hydrolases, utilizing hydrolytic reactions to cleave the mannosidic bonds [8].

Caldanaerobius polysaccharolyticus, a thermophilic anaerobic bacterium isolated from the waste of a corn canning factory in Illinois [14], has been previously shown to possess the enzymatic repertoire required to degrade mannan to fermentable sugars [15-18]. The previously characterized Man5A and Man5B, however, may not be enough to efficiently depolymerize the mannan polysaccharide into monomeric components. Genomic analysis of *C. polysaccharolyticus* revealed the presence of five genes encoding GH130 proteins, with a low overall identity amongst them, likely hinting at different roles in the metabolism of mannan.

The characterization of these five enzymes might provide metabolic insights on how *C. polysaccharolyticus* is able to completely depolymerize mannan into fermentable sugars.

1.2 Objectives

There is still much to learn about how *Caldanaerobius polysaccharolyticus* is able to hydrolyze cellulose and hemicellulose. With only a few enzymes characterized in the mannan utilization pathway, CpMan5A and Man5B alone are not sufficient to efficiently deconstruct mannan into monosaccharides. Furthermore, the presence of five genes encoding the GH130 family of enzymes is intriguing, as differences between the 14

characterized enzymes provide very little information about characteristics shared by proteins in this family.

In an effort to understand the mannan hydrolytic pathway of *C. polysaccharolyticus,* the five GH130 enzymes encoded within the genome of this organism were biochemically characterized. The results reported here identify the differences in the GH130 enzymes of *C. polysaccharolyticus* and provide insights into their mechanisms at the molecular level.

CHAPTER 2: LITERATURE REVIEW

2.1 Plant cell wall polysaccharides

The plant cell wall is composed mainly of polysaccharides such as cellulose, intertwined by hemicelluloses and pectin, as well as the phenolic polymer lignin. It is the most abundant organic compound found in nature and an important energy source for a variety of organisms [19]. Together, the polysaccharides and lignin provide high complexity and rigidity to the plant cell wall [1, 2]. Cellulose, the major constituent of plant cell wall, is comprised of β-1,4-linked D-glucose residues forming linear polymeric chains of about 8,000 – 12,000 glucose units. In its crystalline form, cellulose chains are packed together by hydrogen bonds that form the highly insoluble structures known as microfibrils [8]. Cellulose contains regions within the microfibrils that are noncrystalline in nature, termed amorphous regions [1]. The crystalline regions are highly recalcitrant to degradation by both chemical or enzymatic action, mostly due to how tightly packed the fibers are, but also due to the formation of a water layer on its proximity that hinders accessibility of enzymes or acidic compounds to the fibers [2].

Differently from cellulose, hemicelluloses are heterogeneous in nature and are comprised of a mixture of fiber and sugar components, which vary not only between different plants [19], but also between different tissues within the same plant or even within different cell types and maturity [2]. Hemicelluloses are generally classified based on the main sugar residues comprising the backbone of the structural polymer, and can be divided in three main groups, namely xylan, mannan and xyloglucan. Xylan, the most abundant hemicellulose polymer, found in cereals and hardwood, consists of a

backbone composed of β-1,4-linked D-xylose units. Although cereal xylans display mostly branched structures, with branch frequency and composition varying greatly between polymers from different sources [20], there are cases of linear xylo-polysaccharides found in nature [21, 22]. The xylan backbone can be acetyl-substituted or ramified by an array of different side groups, including but not limited to D-galactose, L-arabinose, glucuronic acid, feruloyl and *p*-coumaroyl residues [23]. Depending on the degree of substitution, the polysaccharide can be described by different names. Cereal xylans are often referred to as arabinoxylans due to the large quantities of L-arabinose ramifications, either in the pyranose or in the furanose forms. Hardwood xylans, which contain large amounts of D-glucuronic acid side chains, are referred to as glucuronoxylans [8].

The backbone of mannan is the most abundant hemicellulose polymer in softwood (Gymnospermae) and in some specialized plant tissues, such as the seeds and fruits of Angiospermae, e.g. ivory nuts and green coffee beans [6, 24]. It is comprised of β-1,4-linked D-mannose units that can form linear chains or substituted by D-glucose to form glucomannan, or branched with D-galactose side chains, forming galactomannan, or sometimes both, forming galactoglucomannan [16, 25]. Mannans can also be utilized as a reserve for non-starch carbohydrates in endosperm walls and vacuoles of seeds and vegetative tissues, and as signaling molecules in plant growth and development [6, 26]. Linear mannans, the simplest of those chains, are homo-polysaccharides containing no more than 5% galactose [26] and are mainly found in ivory nuts and palm oil kernels [27, 28]. The major fraction of hemicellulose in softwoods, glucomannans can represent up to 50% of the hemicellulose content in

coniferous woods [8] and are typically comprised of a 3:1 ratio of β-1,4-linked D-mannose and D-glucose residues [26]. A linear chain of β-1,4-linked D-mannose units with α-1,6-linked D-galactose side chains is the main constituent of galactomannan [8]. Most commonly found in the seeds of plants in the family Leguminoseae, galactomannans have also been identified in other plants such as Ebenaceae and Palmae [29, 30] and their structural characteristics may vary for galactomannans originating from different plant sources [6, 31] (Fig. 2.1A). Galactoglucomannans are comprised of randomly distributed β-1,4-linked D-mannose and D-glucose residues and can possess D-galactose side chains linked to the mannose units [26, 32]. Galactoglucomannans represent another major hemicellulose component of Gymnospermae and can be found in both insoluble and water-soluble forms, the latter being decorated with acetyl residues attached to the main chain [8] (Fig. 2.1B).

The third most commonly found polymer in hemicellulose is xyloglucan, which is comprised of a β-1,4-linked D-glucose backbone with D-xylose substitutions [19]. In this polymer, side chains of L-arabinose and D-galactose substitutions may be linked to the interspersed D-xylose residues in the backbone, and L-fucose residues can be found linked to the D-galactose substitutions. Due to the great diversity of side groups that can be found attached to the main backbone of xyloglucans, they present high structural complexity and variability [1, 33] (Fig. 2.1C).

The most complex component of plant cell walls, pectins are heteropolymers with as many as 17 different monosaccharides organized in an array of distinct polysaccharides and structural groups [34-36]. There are multiple models for their structural organization, the most accepted proposing the distinction in two regions: one

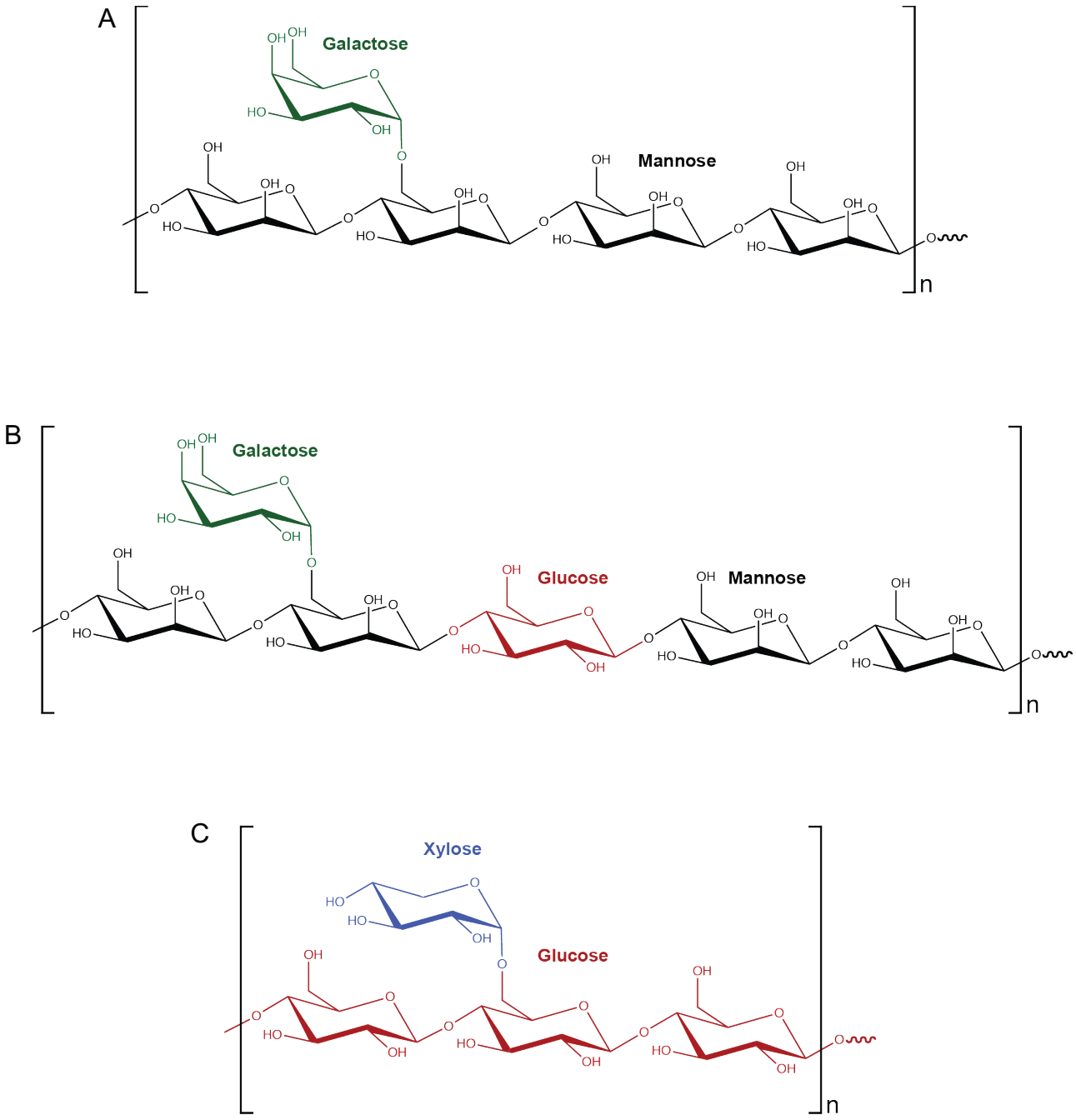


Fig. 2.1 – Schematic representation of prevalent types of hemicellulose
Schematic representation of the structure of (A) galactomannan, (B) glucogalactomannan and (C) xyloglugan. Monomeric sugar units are color coded: mannose is shown in black, glucose in red, galactose in green and xylose in blue.

smooth and one hairy. The “smooth” regions are composed of a backbone of α-1,4-linked D-galacturonic acid (GalA) units, also referred to as homogalacturonans. The GalA residues of the homogalacturonan may be methylated at position *C*-6 or acetylated at positions *O*-2 or *O*-3. They can also be substituted with β-1,3-linked xylopyranose side chains to form the structures referred to as xylogalacturonan [1, 36, 37]. The “hairy" regions, referred to as Rhamnogalacturonan I, exhibit a backbone composed of a single repeating unit of α-1,2-D-galacturonic-L-rhamnose which can be acetylated at the positions *O*-2 and *O*-3 of the GalA residue, or substituted at *O*-4 of the same residue with neutral sugars [8].

Lignin is an aromatic cell wall polymer synthesized by radical coupling of its precursors, coniferyl, sinapyl and *p*-coumaryl alcohol, which form a sequence of guaiacyl (G), syringyl (S) and hydroxypenthyl (H) units [38]. This highly insoluble polymer encrusts and glues together the network of cell wall carbohydrates, stiffening the cell wall structure and granting it rigidity, which is responsible for the protection of plant cells from mechanical stress, osmotic lysis and pathogens [19, 33]. The ratio of G-, S- and H-units differs in softwood, hardwood and grass species. While softwood lignin is mainly composed of G-units, both G- and S-units are abundant in hardwood. In addition to G- and S-units, grasses also have H-units [38].

2.2 Microbial degradation of plant cell wall

Microorganisms have developed, during the course of evolution, physiological mechanisms to cope with a variety of environmental factors and to obtain nutrients from a wide range of sources, such as cellular mechanisms to derive energy from plant

biomass – one of which involves the production and secretion of carbohydrate-active enzymes (CAZymes) [5]. In this context, one of the most common lifestyles of microorganisms is saprophytism, which involves a process of extracellular digestion and processing of dead or decaying organic matter, mainly composed of plant biomass [19]. Saprophytic microorganisms produce multiple enzymes to degrade plant cell materials, and the enzyme groups that act on the same substrate type are known as enzyme systems [39].

For microorganisms to hydrolyze and metabolize insoluble cellulose, extracellular cellulases must be produced that are either free or cell associated. Components of cellulase systems were first classified based on their mode of catalytic action and have more recently been classified based on structural properties [40].

Although cellulose is a homopolymer consisting of linear chains of β-1,4-linked glucose units, its complete enzyme catalyzed degradation to glucose requires the concerted action of three different enzymes. First, endo-1,4-β-glucanases (EG) (EC 3.2.1.4) randomly cleave internal amorphous sites in the polysaccharide chain, generating oligosaccharides of various lengths and new chain ends. Exoglucanases (EC 3.2.1.91) act on the reducing or non-reducing ends of the cellulose and oligosaccharides from endoglucanase activity, liberating cellobiose or glucose. Finally, β-glucosidases (BG) (EC 3.2.1.21) and cellodextrinases act by breaking down the cellobiose and other cellodextrins into glucose monomers. Exoglucanases have also been described to act on microcrystalline cellulose, presumably peeling cellulose chains from the microcrystalline structure [5, 41].

Hemicellulose, being a structurally complex heteropolymer, requires the action of a wide array of depolymerizing and debranching enzymes. The presence and degree of substitutions and acetylation links are known to slow the action of the depolymerizing enzyme activities by physically restricting the access of enzymes to the β-1,4-mannose linkages in the mannan backbone [6, 24, 26, 27, 42], as is also the case for di-ferulated cross links in xylan [1, 3, 33] backbones. Examples of enzymes involved in the debranching activities include α-L-arabinofuranosidase (EC 3.2.1.55), α-glucuronidase (EC 3.2.1.139), acyl- and acetyl xylan esterases (EC 3.1.1.6 and EC 3.1.1.72, respectively), α-galactosidases (EC 3.2.1.22) and feruloyl esterases (EC 3.1.1.73). Enzyme activities involved in depolymerizing the backbone include endo-1,4-β-xylanases (EC 3.2.1.8) and β-xylosidases (EC 3.2.1.37) for xylan backbones [1, 3, 33, 42] and endo-1,4-β-mannanases (EC 3.2.1.78) and β-mannosidases (EC 3.2.1.25) for mannan backbones [6, 26, 42].

2.3 Industrial applications of cell wall degradation

Bioconversion of vegetal biomass, especially hemicellulose and lignin, has increasingly received more attention in recent years. This effect might be a consequence of the increasing interest in producing cheaper, renewable energy via second generation ethanol [2, 43-45] or other biofuels [46-49], or to convert the large amounts of agricultural and industrial biomass waste into commodities by utilizing biorefineries [24, 50, 51]. Biological processes that are capable of converting lignocellulosic biomass into valuable products are sought by an increasing number of companies, such as Archer Daniels Midland Corporation, Cargill, Dow Chemical

Company, Huntsman Corporation and Solvay [52, 53]. Processes to make bio-sourced intermediate chemicals, such as acrylic acid, could be implemented to existing chemical producing facilities as intermediates for the production of higher value-added chemicals, such as acrylic esters, without requiring expensive changes to the current manufacturing pipelines [8].

There is a growing interest in processing byproducts from the food industry feedstock. For example, for wheat and barley, important products in the food industry in Europe, the valorization of the processing byproducts has become vital for securing the future economy [54-56]. Residue disposal brings environmental concerns and a prospected increase in cost that are offset by the revenue generated by the pipeline upgrade [8]. The same holds true for agricultural residues, such as from corn crops, with seeds consisting of about 60% starch and applicability for animal feed or ethanol production, or for agro-industrial residues such as in the sugar and ethanol production from sugarcane in Brazil [43, 44, 57, 58]. Today, most of these residues are burnt in boilers that generate energy to fuel the extraction processes [58, 59]. However, converting the biomass in these residues to ethanol will equate to an increase in production in the order of 125 L of ethanol per ton of cultivated sugarcane, or roughly 47% of the amount obtained from first generation ethanol production [60].

Enzymes with the ability to degrade hemicellulose and lignin also possess applicability in industries other than the production of biofuels, as is the case of the paper industry. Alkaline β-mannanases and β-xylanases are essential in the enzymatic bleaching of softwood pulps as this process helps reduce environmental impact by chlorine consumption [27, 61]. Since the majority of pulps are derived from softwood,

with 15 to 20% of their hemicellulose contents comprised of galactomannans, mannanases with high substrate specificity would allow for less cellulose degradation and a higher quality paper [8]. Pulping, as most industrial processes involving enzymatic degradation of vegetal cell wall components, is carried out at elevated temperatures; therefore, thermophilic hemicellulases offer significant advantages over their mesophilic counterparts in regards to industrial application. Thermophilic enzymes are also more desirable as they remain active and retain efficiency without incurring in process cooling expenses [8].

2.4 *Caldanaerobius polysaccharolyticus*

Caldanaerobius polysaccharolyticus, originally *Thermoanaerobacterium polysaccharolyticus* [14], is a straight rod-like, motile (flagellate), Gram-positive bacterium that occurs singly or in pairs. Although the organism is aerotolerant, growth occurs under anaerobic conditions, at temperatures above 45˚C and up to 70˚C, with the best observed growth temperature ranging between 65 – 68˚C and a doubling time at 68˚C, with glucose as the carbon source, of 2.1 hours. The optimum pH for growth ranges between 6.8 – 7.0. Spore formation has not been observed for this organism. This organism is able to utilize glucose, arabinose, galactose, lactose, maltose, mannose, rhamnose, salicin, sucrose, trehalose, xylose, cellobiose, raffinose, melibiose, melizitose and pyruvate. The end products of fermentation on a minimal medium with 0.5 % (w/v) glucose, as the carbon source, are ethanol (30 mM), acetate (20 mM), formate (11 mM) and lactate (5 mM) [8].

The bacterium *C. polysaccharolyticus* was isolated from organic waste from a canning factory situated in Hoopeston, Illinois, USA, which seasonally processes sweet corn and other vegetables. The canning waste was characterized by active gas evolution, low pH (5.5), high concentrations of short-chain fatty acids (31.6 mM acetate, 8.2 mM propionate, 26.0 mM butyrate and 6.4 mM valerate) as well as heat generation [8].

Based on metabolic differences and phylogenetic analysis, *T. polysaccharolyticus* and two other bacteria were grouped as a new genus, *Caldanaerobius* sp., which was placed into the order '*Thermobacteriales*' [62] in the class *Clostridia* of the phylum *Firmicutes*. The genus consists of anaerobic and thermophilic chemo-organotrophs that do not require yeast extract for growth and produce formate from glucose fermentation, unlike other members of the genus *Thermoanaerobacterium*. The G+C content of the genomic DNA for *Caldanaerobius* spp. is 37 – 46 mol%, which is also higher than that of the genus *Thermoanaerobacterium*, which are reported to have a genomic G+C content below 36 mol%. The type species for the genus *Caldanaerobius* is *Caldanaerobius fijiensis* [63].

2.5 Mannan utilization by *C. polysaccharolyticus*

Caldanaerobius polysaccharolyticus has been observed to grow on glucose and mannose, and to depolymerize mannan polysaccharides into shorter manno-oligosaccharides (MOS) and to mannose units [14-16]. The analysis of the genome of this bacterium revealed the presence of two genes encoding glycoside hydrolase family 5 (GH5) proteins, named Man5A and Man5B [8]. The Man5A of *C. polysaccharolyticus*

consists of N-terminal signal-peptide, followed by a GH5 catalytic domain that is flanked at the C-terminus by a sequence of unknown function. The region of unknown function is followed by two family 16 carbohydrate binding modules (CBM16) in tandem. The tandem CBMs are followed by three surface layer homology repeats (SLH) that are predicted to anchor Man5A to the cell surface. Man5B, on the other hand, is composed of a single-GH5 module and lacks a signal peptide [15, 16]. The domain architecture, described above, suggests that Man5A is an extracellular endo-mannanase capable of degrading mannan and carboxymethyl cellulose (CMC) [15] and larger chain oligosaccharides, but with minimal activity against *p*NP-linked substrates and oligosaccharides with a degree of polymerization (DP) lower than four [8]. The genomic context of the cluster containing Man5A revealed that the gene is preceded by a putative transporter, similar to the membrane component of an ABC transporter. This transporter-like polypeptide is likely involved in the transport of monomers and short oligomers derived from the hydrolysis of polysaccharides by the enzyme [8]. This prediction is supported by the lack of activity on *p*NP-linked substrates.

The second GH5 protein, Man5B, is likely an intracellular protein, based on the absence of a putative signal peptide. This enzyme displayed hydrolytic activity on *p*NP-linked β-D-mannopyranoside and β-D-cellobioside, in addition to oligosaccharides composed of either mannose or glucose. The optimal temperature was slightly higher for activity on the β-1,4-glucosidic links in comparison with the β-1,4-mannosidic [8], a behavior observed also for Man5A [8]. The end products after a 12h incubation of Man5B, at its optimal pH and temperature, with MOS of varying lengths were a mixture of mannobiose and mannose. The enzyme was unable to carry out the degradation of

cello-oligosaccharides to completion after the same incubation time, and incubation with oligosaccharides of varying lengths resulted in a mixture of smaller products [8].

The genes encoding the predicted extracellular Man5A, which may be responsible for the initial hydrolysis of the mannan polysaccharide, and Man5B, predicted to function intracellularly, are located in separate gene clusters on the *C. polysaccharolyticus* genome [18]. This strategy is in agreement with a gene cluster that encodes proteins involved in xylan degradation, also described in *C. polysaccharolyticus*, where the gene encoding the predicted extracellular endoxylanase is located outside of the xylo-oligosaccharide metabolism gene cluster [18, 64].

The complete degradation of mannan into fermentable sugars require the concerted effort of multiple enzymatic activities, including endo-mannanases to cleave the polysaccharide backbone. The oligosaccharides generated are then attacked by β-mannosidases and β-glucosidases to produce the smaller oligosaccharides into monomeric units of either mannose or glucose. Alpha-galactosidases and acetyl-esterases work in synergy with these enzymes to remove the branching units (Fig. 2.2). It is predicted that in *C. polysaccharolyticus*, endo-mannanase activity derives from Man5A, and the degradation of the oligosaccharides, to some extent, is performed by Man5B. In combination, these two enzymes would allow for the degradation of most of the mannan polysaccharides to a mixture of mannobiose and mannose. Moreover, the presence of the putative membrane-bound ABC transporters reported in their gene clusters would account for the import of oligosaccharides from Man5A activity into the cytoplasm, where they can be further metabolized by Man5B and other enzymes [8].

2.6 GH130 family enzymes

The GH130 family (in the CAZy – Carbohydrate Active Enzyme-Database) was created based on the identification of 4-O-β-D-mannosyl-D-glucose phosphorylase activity (EC 2.4.1.281) for a protein derived from the gene BF0772 of *Bacteroides fragilis* [9]. The GH130 family consists of enzymes that catalyze the phosphorolysis of β-mannosidic linkages at the non-reducing end of substrates, although these enzymes are likely involved in degradation of β-1,4-D-mannosyl-N-acetyl-D-glucosamine linkages in the core of N-glycans [11]. Other activities described for the family include β-1,4-mannooligosaccharide phosphorylase (from *Ruminococcus albus*) [10], 1,4-β-mannosyl-N-acetylglucosamine phosphorylase (unknown human gut bacterium mannoside phosphorylase – UhgbMP) [8], 1,2-β-oligomannan phosphorylase and β-1,2-mannobiose phosphorylase (both from *Thermoanaerobacter* sp. X-514) [12].

The three-dimensional structures of a few GH130 family members have been reported. The structure of *Bacteroides fragilis* BfMGP has been reported in complex with an array of substrates and/or products, as listed: phosphate; 4-O-β-D-mannosyl-D-glucose and phosphate; mannose, glucose, and phosphate; and α-mannose-1-phosphate [13]. This enzyme forms a homohexamer where each monomeric unit possesses a five-bladed β-propeller fold. Long α-helices were observed at the N- and C-termini and are predicted to be responsible for the quaternary structure formation [13].

Phylogenetic analysis of GH130 proteins pertaining to microorganisms from the human gut allowed for the proposition of two subfamilies, namely GH130_1 and

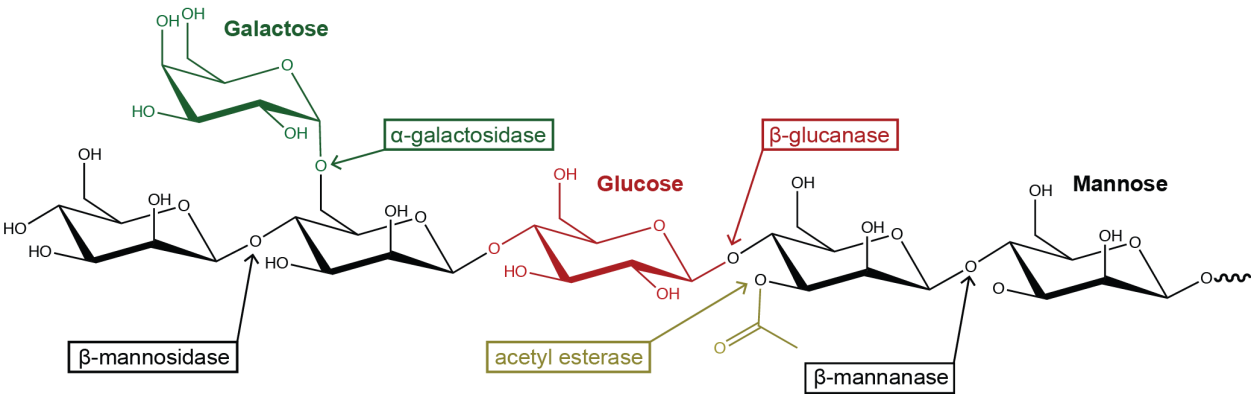


Fig. 2.2 – Schematic representation of acetyl-galactoglucomannan
Schematic representation of the structure of acetyl-galactoglucomannan showing a glucose interspersion (red) and the possible galactose (green) and acetyl (ocre) substitutions. Enzymatic activities required for its degradation and the linkages they attack are evidenced.

GH130_2. Moreover, a large number of the GH130 protein sequences do not share enough characteristics with either subfamily, especially in the phosphate binding site. These GH130 proteins differing in the phosphate binding site have been designated as non-clustering (GH130_NC) due to the lack of information about their activities [7]. Both subfamilies GH130_1 and GH130_2 seemingly share several amino acids in the active site, including a conserved phosphate binding site consisting of a tyrosine, a histidine, two arginines and a fifth variable charged amino acid (Y103, H174, R150, R168 and N151 in UhgbMP [8]; Y284, H273, R191, R209 and K254 in BfMGP [8]). The two subfamilies appear to differ in substrate specificity due to variations in the amino acid residues in their carbohydrate binding sites.

Proteins in the GH130_1 family have been described to perform the phosphorolysis of β-1,4-mannosyl glucose at the non-reducing end of substrates with net inversion of anomeric configuration and thus yielding an α-mannose-1-phosphate. This activity has been demonstrated for a 4-O-β-D-mannosyl-D-glucose phosphorylase of *Bacteroides fragilis* (BfMGP) that produces α-mannose-1-phosphate and glucose from 4-O-β-D-mannosyl-D-glucose and inorganic phosphate as substrates [9] (Fig. 2.3A). The mechanism for GH130_1 activity has been proposed to involve a catalytic Aspartate that donates a proton to position *O*3 of the mannosyl group bound to sub-site -1 [65] and a second proton that is intramolecularly transferred to the glycosidic oxygen from the hydroxyl group at position 3. In other inverting phosphorylases, a general acid catalyst is proposed to directly protonate the glycosidic oxygen [13]. In contrast, subfamily GH130_2 has been described to perform a similar reaction, but with high

specificity for β-1,4-mannosidic bonds and much lower affinity for substrates substituted with glucose than GH130_1 [7, 10, 66].

A comparison between 369 protein sequences made it possible to select D104, E273, and D304 of the GH130_2 subfamily exemplar UhgbMP as the putative catalytic amino acid residues for the family (Fig. 2.3B), as mutations in any of these acidic amino acid residues largely reduced or completely abolished the enzymatic activity [7]. The three-dimensional structure analysis demonstrated that only D131 of BfMGP, corresponding to D104 of UhgbMP, is situated near the glycosidic oxygen, which is in agreement with what had been postulated for UhgbMP, although the aspartate appeared to be too distant from the glycosidic oxygen for direct protonation. Hence, a unique proton relay mechanism for GH130 enzymes was proposed on the basis of the three-dimensional structure of BfMGP [13].

The group of non-clustering GH130_NC encompasses the largest number of enzymes currently assigned to this family and to date, it remains unclear what are the optimal substrates for these enzymes. The absence of the inorganic phosphate (P_i) binding residues in some sequences within this group hints at the presence of hydrolytic activity in these enzymes. This hypothesis is supported by the presence of a glutamic acid residue replacing a key tyrosine in exemplars of GH130_NC exhibiting activity towards β-1,2-linked oligomannan [7, 67, 68], which likely plays a role similar to that of a general base catalyst in hydrolytic enzymes (Fig. 2.3C). Some enzymes within the GH130_NC group possess the conserved phosphate-binding amino acid residues [12, 67, 69] and exhibit activity towards β-1,2-linkages (Fig. 2.3D).

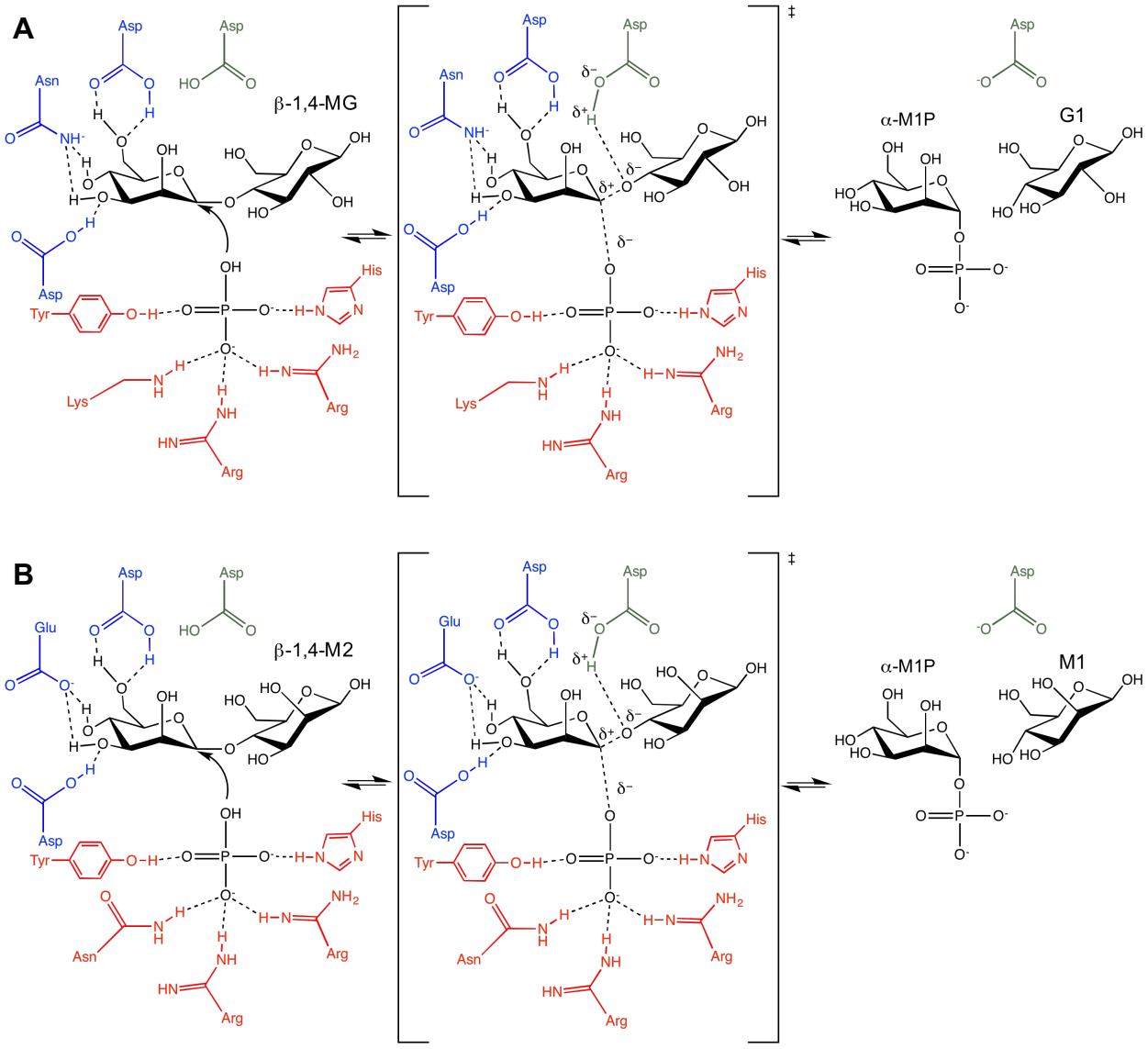


Fig. 2.3 – Schematic mechanisms proposed for the GH130 family enzymes
(A) General mechanism proposed for GH130_1, consisting of a single displacement. Five conserved amino acids (shown in red) coordinate the phosphate molecule, while other three (shown in blue) coordinate the mannose (M1) at the non-reducing end and an aspartate residue (shown in green) donates a proton to help disrupt the β-1,4-linkage and results in an inversion of anomeric configuration at the C1 position. Glucose (G1) and α-mannose-1-phosphate (M1P) are released as the products when the reaction occurs on mannosyl-glucose (β-1,4-MG). (B) General mechanism proposed for GH130_2, consisting of a single displacement. Five conserved amino acids (shown in red) coordinate the phosphate molecule, while other three (shown in blue) coordinate the mannose (M1) at the non-reducing end and an aspartate residue (shown in green) donates a proton to help disrupt the β-1,4-linkage and results in an inversion of anomeric configuration at the C1 position. Mannose and α-mannose-1-phosphate (M1P) are released as the products when the reaction occurs on mannobiose (M2). (C) General mechanism utilized by the GH130_NC with a conserved phosphate binding site and activity on β-1,2-linked manno-oligosaccharides. (D) General mechanism utilized by the GH130_NC lacking a phosphate binding site, with hydrolytic activity on β-1,2-linked manno-oligosaccharides.

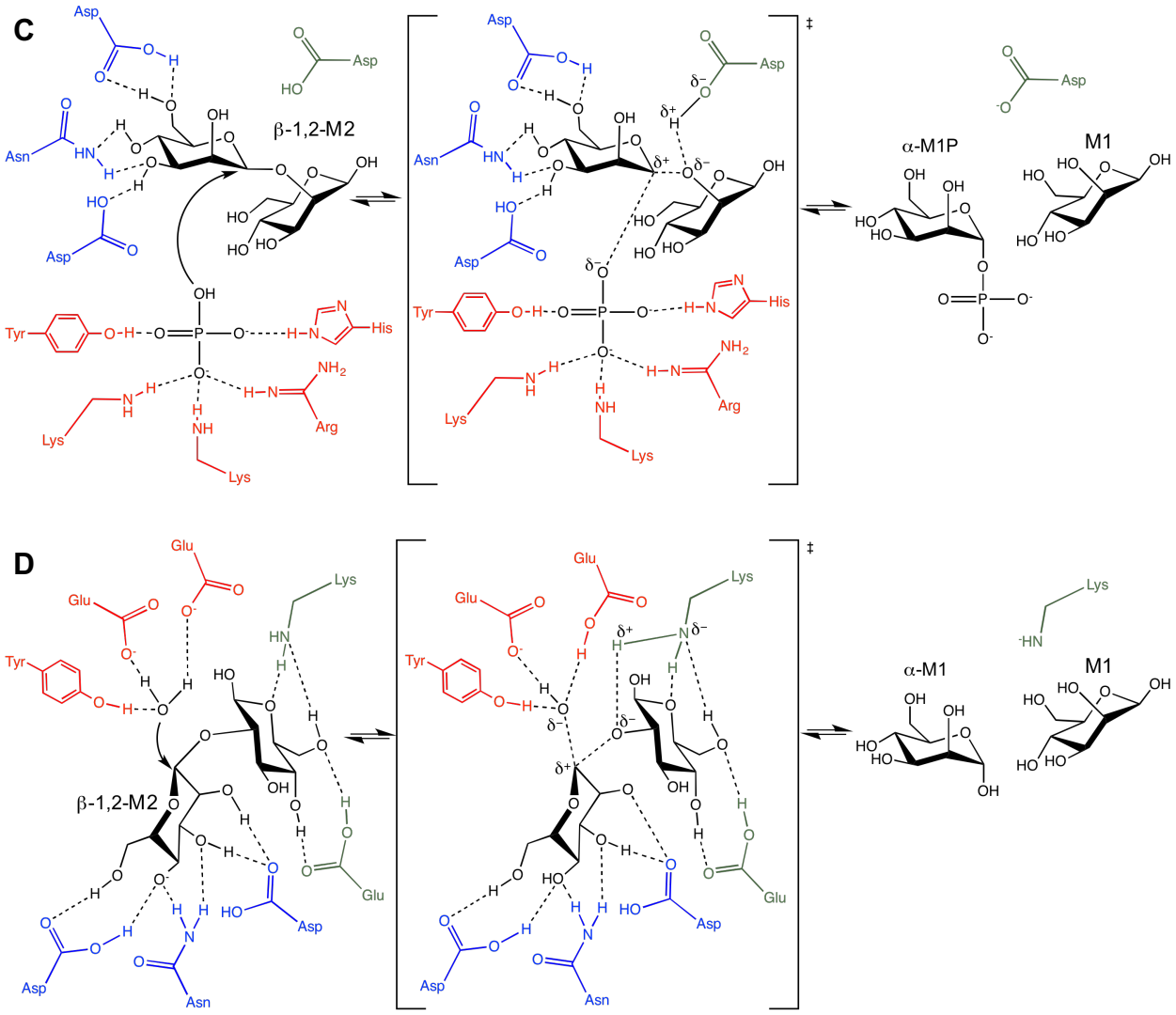


Fig. 2.3 - continued.

CHAPTER 3: METHODOLOGY

3.1 Bacterial strain and materials

The cells of *C. polysaccharolyticus* ATCC BAA-17 was obtained from the Department of Animal Sciences, University of Illinois at Urbana-Champaign [14]. *Escherichia coli* JM109, DH5α and BL21-CodonPlus (DE3) RIPL competent cells were purchased from Stratagene (La Jolla, CA). The pET-46 Ek/LIC vector kit was obtained from Novagen (San Diego, CA) and QIAprep Spin Miniprep kit was purchased from Qiagen (Valencia, CA). Oligosaccharides α-mannose-1-phosphate (M1P), mannose (M1), mannobiose (M2), mannotriose (M3), mannotetraose (M4), mannopentaose (M5), mannohexaose (M6), glucose (G1), cellobiose (G2), cellotriose (G3), cellotetraose (G4), cellopentaose (G5), cellohexaose (G6), arabinose (A1), arabinobiose (A2), arabinotriose (A3), arabinotetraose (A4), arabinopentaose (A5), arabinohexaose (A6), α-1,6-galactosyl-mannotriose (GaM3), and α-1,6-galactosyl-mannobiose mixed with mannotriose (GaM2+M3) were purchased from Megazyme (Bray, Ireland) or Sigma-Aldrich (St. Louis, MO). All other reagents were of the highest possible purity and were purchased from Sigma-Aldrich (St. Louis, MO) or Thermo Fisher Scientific (Waltham, MA), unless otherwise stated.

3.2 Gene cloning[2]

Analysis of the genome of *Caldanaerobius polysaccharolyticus*, sequenced by the Joint Genome Institute (JGI, Walnut Creek, CA), revealed the presence of a total of

[2] Genome analysis performed by Inhyuk Kwon, cloning performed by Inhyuk Kwon and Hans Müller Paul

five putative GH130 genes. Each of the genes encoding the five GH130-family enzymes (CALPO_RS0106045, CALPO_RS0100380, CALPO_RS0109450, CALPO_RS0109445 and CALPO_RS0110910) was amplified from the genomic DNA of *Caldanaerobius polysaccharolyticus* ATCC BAA-17 via PCR using Phusion® high-fidelity DNA polymerase (New England Biolabs, Ipswich, MA) and the primers listed in Table 3.1. The PCR products were purified and cloned into the pET-46b Ek/LIC vector (Novagen, San Diego, CA). Plasmids were constructed according to the manufacturer's instruction and transformed into competent *E. coli* DH5α cells using the heat-shock method. Colonies were picked from Lysogeny broth (LB) agar plates containing 100 μg/ml of ampicillin and cultured in liquid LB medium containing the antibiotic at the same concentration at 37 ºC with shaking at 200rpm overnight. Each of the recombinant plasmids was purified from the respective culture using a QIAprep Spin Miniprep kit (Qiagen, Hilden, Germany). The DNA insert of each plasmid was sequenced to confirm the integrity of the cloned genes (W. M. Keck Center for Comparative and Functional Genomics at the University of Illinois at Urbana-Champaign).

3.3 Recombinant protein expression and purification

The recombinant plasmids were transformed into *E. coli* BL21-CodonPlus (DE3) RIPL competent cells using the heat-shock method, and cells were incubated overnight at 37 ºC on selective LB agar plates supplemented with 100 μg/ml of ampicillin and 50 μg/ml of chloramphenicol. Individual colonies harboring each recombinant plasmid were picked and inoculated in 10 ml fresh LB medium containing the antibiotics at the same

Table 3.1 – Primers used in cloning of GH130 genes

Gene	Primer	Sequence
CpMan130A	GH130-2 F	5'- GACGACGACAAGATGAGCAAACAGTTGATAGTCGGGG -3'
	GH130-2 R	5'- GAGGAGAAGCCCGGTCACAGAGAGTTATTGGATTTTACA -3'
CpMan130B	GH130-1 F	5'- GACGACGACAAGATGAGCAAACAGTTAATAGTCGGGGAG -3'
	GH130-1 R	5'- GAGGAGAAGCCCGGTTAACTATTGGACTTTATAAAGTCTATCACTTCGTCC -3'
CpMan130C	GH130-3 F	5'- GACGACGACAAGATGTTTAAGCTCAAGAGGTTATCAGATAAACCGGTATTATCTC -3'
	GH130-3 R	5'- GAGGAGAAGCCCGGTTAAAACTTCACTTTGCTTTTTTCTACAGCGGCCAC -3'
CpMan130D	GH130-4 F	5'- GACGACGACAAGATGATAAAGTTAGAGAGATTGACGGATAAGCCGGTTTTAC -3'
	GH130-4 R	5'- GAGGAGAAGCCCGGTTAATCAAAAGTTATATCCTTCATATCTATACCCGCAACACC -3'
CpMan130E	GH130-5 F	5'- GACGACGACAAGAATGAACAGTTACATAGACCCGTCT -3'
	GH130-5 R	5'- GAGGAGAAGCCCGGTTAGCTTTCTTTTTGCCCAGGACA -3'

concentrations, and incubated at 37 ºC and 200 rpm for 6h. The cultures were then transferred into a 2.8 liter Erlenmeyer (Fernbach design) flask containing 1 liter LB medium supplemented with antibiotics and cultured at 37 °C until the optical density at 600 nm (OD_{600}) reached 0.6. At that point, isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 0.1 mM and the temperature was lowered to 16 ºC and culturing continued for another 16h.

For experiments that required higher protein concentrations, pre-cultures were carried out in 15 ml fresh LB medium containing 100 µg/ml of ampicillin and 50 µg/ml of chloramphenicol and incubated at 37 ºC with shaking at 200 rpm for 6h. Each 15 ml pre-culture was transferred to 1.5 liter cultures in BioFlo 310 Bioreactors (New Brunswick Scientific), with selective SB media (Difco™ Select APS™ Super Broth, prepared following the manufacturer's orientations) containing 100 µg/ml of ampicillin and 50 µg/ml of chloramphenicol. The cell culturing was continued for 8h at 37 ºC. Recombinant gene expression was induced by adding isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.25 mM. The temperature was lowered to 20 ºC and the cell culturing was continued for another 16h. Gas flow was kept at 4 SLPM (standard liters per minute) and agitation was kept at 200 rpm throughout the entire process.

Cells were harvested by centrifugation (4,000 rpm, 30 min, 4ºC) and the cell pellets were re-suspended in 25 ml of ice-cold lysis buffer and lysed using an EmulsiFlex C3 Homogenizer, from Avestin (Ottawa, ON, Canada). The cell debris were separated by centrifugation (10,000 rpm, 30 min, 4°C), and the supernatants were incubated at 65 ºC for 30 minutes to denature the mesophilic proteins from *E. coli*.

Centrifugation under the same conditions was performed again to remove the denatured *E. coli* proteins. The clarified supernatants were applied to Talon® metal affinity resin (Clontech Laboratories, Mountain-View, CA) that had been already equilibrated with binding buffer (100 mM HEPES, 500 mM NaCl, 0.5 mM TCEP, 10% glycerol, pH 8.0). Purification was performed according to the protocol from the manufacturer and the utilized buffers are listed in Table A.1 (Appendix A). The sizes of all purified proteins were confirmed by SDS-PAGE stained with Coomassie brilliant blue G-250. The purified proteins were concentrated and the buffers exchanged to a storage buffer using Spin-X® UF 20 (10k MWCO PES) concentrators from Corning Life Sciences (Corning, NY). Recombinant proteins were stored at -20 ℃ until used. The concentration of each purified protein was determined using the NanoDrop 1000 apparatus from Thermo Fisher Scientific Inc. (Waltham, MA), based on their molecular weight and extinction coefficients (table 3.2).

3.4 Structure determination of CpMan130B[3]

CpMan130B was purified as described above but with an additional size exclusion chromatographic step (Superdex™ 200 Hiload™ 16/60 size exclusion column, GE Healthcare, Piscataway, NJ) using a final buffer composed of 20 mM HEPES pH 7.5, 100 mM KCl. Initial crystallization conditions were obtained by sparse-matrix sampling method using commercial screens. Crystals of the apoprotein were grown using the hanging vapor drop diffusion method. One microliter of the protein was

[3] Protein expression and purification for this experiment were performed by Inhyuk Kwon. Structure determination was performed by Jonathan R. Chekan from Dr. Satish Nair's group.

Table 3.2 – Biochemical characteristics of the five GH130 proteins

Characteristic	CpMan130A	CpMan130B	CpMan130C	CpMan130D	CpMan130E
Number of amino acids	329	327	297	302	317
M.W. (kDa)	37.5	37.3	34.2	34.4	36.0
Extinction coefficient	73,800	73,800	68,300	77,350	69,330
Theoretical pI	5.61	5.61	5.55	5.37	6.68

mixed with 1 µl of precipitant solution and equilibrated over a well containing the precipitant solution at 9 ºC. Crystals grew within 2 weeks and were briefly soaked in precipitant solution supplemented with 10% ethylene glycol prior to vitrification by direct immersion in liquid nitrogen.

3.5 Isothermal Titration Calorimetry (ITC)

Isothermal Titration Calorimetry (ITC) was used to assess the binding capability of CpMan130B, a representative GH130_2 enzyme, to phosphate, mannobiose, cellobiose and combinations of phosphate with mannobiose and phosphate with cellobiose. Measurements were carried out at 25 ºC using a VP-ITC calorimeter (MicroCal Inc., Northampton, MA) following the manufacturer's instructions. Enzyme samples were extensively dialyzed against a 50 mM HEPES buffer, pH 6.5, and the ligands were prepared as solutions in the buffer obtained after the dialysis was completed. The protein sample at 50 µM was injected with 28 successive 10 µl aliquots of the ligands or ligand combinations, as presented in Table 3.3, at intervals of 300 seconds [70].

3.6 Determination of optimal pH and temperature for activity

The determination of optimal pH and temperature was carried out by analyzing the effects of pH and temperature on mannobiose hydrolysis. Two independent variable optimization was carried out for each enzyme, following the same experimental configuration (90nM enzyme and 250µM mannobiose, 10 min incubation). In the first stage, the reactions were conducted in 0.1 M of two different buffers, citrate-phosphate,

Table 3.3 – Experimental conditions for ITC

Tested ligand	Ligand concentration (mM)		
	Phosphate	Mannobiose	Cellobiose
Phosphate	2.5	0	0
Mannobiose	0	1	0
Cellobiose	0	0	1
Phosphate + Mannobiose	2.5	1	0
Phosphate + Cellobiose	2.5	0	1

with pH ranging from 3.0 to 7.0, and sodium phosphate, with pH ranging from 6.0 to 8.0, at 65 ºC – based on the culturing temperature for the organism –, to allow the identification of the optimum pH for each enzyme. In the second stage, each enzyme was analyzed for its temperature optimum by conducting the hydrolytic reactions at their individual optimal pH and temperature varied from 35 to 85 ºC with 5 ºC increments. Samples from both assays were analyzed by high-performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD) utilizing a ICS-5000+ high-performance liquid chromatograph (HPLC) system equipped with a CarboPac PA10 column (Dionex Corp., Sunnyvale, CA).

3.7 Enzymatic activity of GH130-family enzymes on different substrates

To assess the activity of the GH130-family enzymes on several oligosaccharides, a set of reactions was carried out for each enzyme in 100 mM citrate-phosphate buffer, at a final enzyme concentration of 90 nM and a final substrate concentration of 250μM. Reactions were carried out under optimal pH and temperature conditions specific to the enzyme. Hydrolytic activities were carried out in triplicates together with a negative control experiment (no enzyme added to reaction mixture). The substrates tested included mannobiose (M2), mannotriose (M3), mannotetraose (M4), mannopentaose (M5), and mannohexaose (M6). Samples were collected after three different incubation times (0, 1 hour and 12 hours). The end products of each reaction were then analyzed by high-performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD) on a ICS-5000+ high-performance liquid chromatograph (HPLC) system equipped with a CarboPac PA10 column (Dionex Corp, Sunnyvale, CA).

3.8 Complementarity with enzymes in the mannan degradation pathway

Each of the five GH130 enzymes from *C. polysaccharolyticus* was incubated with either the *C. polysaccharolyticus* endomannanase (Man5A) or the β-mannosidase (Man5B) or both enzymes. The hydrolytic reactions were conducted in the same phosphate-containing buffer as described above. Reactions were carried out with β-1,4-mannan as the substrate. The end products after twelve hours of incubation for each reaction were then analyzed by high-performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD) on a ICS-5000+ high-performance liquid chromatograph (HPLC) system equipped with a CarboPac PA10 column (Dionex Corp, Sunnyvale, CA).

3.9 Analytical gel filtration for determination of molecular weight

For each GH130-family protein cloned from *C. polysaccharolyticus*, a 1 ml sample of purified protein at a concentration of approximately 2 mg/ml in 50mM HEPES pH 7.5, 300 mM NaCl buffer was loaded onto a HiLoad™ 16/60 Superdex™ 200 column using an ÄKTAxpress fast protein liquid chromatography (FPLC) system (GE Healthcare, Piscataway, NJ). The column was equilibrated with a 0.22 μm filtered (Express™PLUS, Millipore, Billerica, MA) buffer composed of 20mM HEPES pH 7.5, 100 mM NaCl. Chromatography was performed at a flow rate of 1 ml/min at 4 ºC. The elution volumes were compared to a standard curve obtained by utilizing the same method and a gel filtration standard containing a mixture of thyroglobulin, bovine γ-globulin, chicken ovalbumin, equine myoglobin, and vitamin B_{12}, with molecular weights

ranging between 1,350 to 670,000 daltons (Bio-Rad, Hercules, CA). Data were collected using the UNICORN v5.31 software (GE Healthcare, Piscataway, NJ) and analyzed using Prism 5.0a (GraphPad Software Inc., La Jolla, CA).

3.10 Amino acid sequence phylogeny and homology modeling[4]

Multiple sequence alignment of all 951 GH130 amino acid sequences available in the CAZy database (May 2016) and the five GH130 amino acid sequences obtained from *C. polysaccharolyticus* was performed with MUSCLE. A distance matrix was generated from the alignment file and exported using Geneious version 9.1.4 [8]. The alignment was then processed utilizing R version 3.3.0 [71] and the 'rattle' package [72] and the output result file was subjected to hierarchical clustering using Ward's method, as previously described [8]. The phylogeny was assembled utilizing the FastTree 2.1 approximately maximum-likelihood algorithm and the resulting tree was visualized with Dendroscope 3 [8]. The phylogenetic tree was used to examine whether each GH130 protein in *C. polysaccharolyticus* clustered with a proposed GH130 subfamily (GH130_1, GH130_2 or _NC) [7, 66]. A detailed methodology, including the codes utilized in R, can be seen in Appendix B.

A protein structure model for the five GH130_2 Man130B from *C. polysaccharolyticus* was generated using the MODELLER software (v9.16) [73, 74]. A single template approach was used. Here, the amino acid sequence of the protein which will be modeled was aligned with a template. The template is a protein with a

[4] Protein structure modeling was performed by Rafael C. Bernardi at the Beckman Institute for Advanced Science and Technology in the University of Illinois at Urbana-Champaign. Amino acid sequence alignment and phylogeny assembly were performed and analyzed by Celia Mendez-Garcia and Hans Müller Paul.

known crystal structure. For this experiment, the unknown human gut bacteria mannoside phosphorylase (UhgbMP, PDB: 4UDI) was utilized as a template. This protein was chosen because it was the first described GH130_2 protein to have its crystal structure deposited to the Protein DataBank (PDB) [75]. This protein also had co-crystals with phosphate and either mannobiose or cellobiose. Quality of the models was evaluated by comparison to the crystal structure obtained experimentally.

The protein structure models were analyzed, using the software Visual Molecular Dynamics (VMD, v. 1.9.2) [76], in order to observe the inorganic phosphate and substrate molecules in the active site. The models were examined for insights into the amino acids involved in the substrate binding, phosphate coordination and overall enzymatic activity.

3.11 NMR spectroscopy for end products identification[5]

Synthetic reactions were conducted for the five GH130 enzymes by incubating each enzyme with a mixture of mannose and α-mannose-1-phosphate (250µg each) in 600 µL of deuterium oxide (D_2O), at 65 ºC for 12h. Enzymes were inactivated by incubating at 100 ºC for 10 minutes. Samples taken from the reaction products were prepared by adding a trace amount of Acetone for referencing the chemical shift [77]. One-dimensional ^1H NMR analysis was performed and, after one step of water suppression, homonuclear correlation spectroscopy (COSY) and heteronuclear single-quantum correlation spectroscopy (HSQC) were performed to correlate the Hydrogen and Carbon atoms in the reaction products. Experiments were performed in an Agilent

[5] NMR data analysis was performed by Daniel Wefers and Hans Müller Paul.

600MHz NMR[6] with a 14.1 Tesla 54 mm bore Agilent Premium Compact Shield Superconducting Magnet and an HCN Probe ($^{1}H\{^{13}C/^{15}N\}$ 5mm PFG Triple Res Probe), in the IGB Core Facility.

[6] The 600MHz NMR in the IGB Core is funded by NIH grant number S10-RR028833.

CHAPTER 4: RESULTS

4.1 Five GH130 genes from *C. polysaccharolyticus* ATCC BAA-17

Caldanaerobius polysaccharolyticus has been previously shown to be able to degrade and grow on mannan polysaccharides [14-16]. Moreover, this thermophilic bacterium has been shown to up-regulate genes in two mannanase-containing clusters when grown on glucomannan and galactomannan as the primary carbon sources [18]. Both clusters containing the mannan-degradation enzymes Man5A and Man5B also contain genes encoding proteins in the recently described glycoside hydrolase family 130 (GH130). The two genes were annotated in the genome sequenced by the Joint Genome Institute (JGI) as CALPO_RS0100380 and CALPO_RS0106045 and are hereon referred to as Man130A and Man130B, respectively. Analysis of the genome sequence revealed the presence of three other putative GH130 encoding genes. The three genes were located outside of the mannan-degradation gene clusters of *C. polysaccharolyticus* (Fig. 4.1). It is important to note that besides the two genes contained within the previously described mannan-degradation clusters, there is a third cluster containing a putative transcriptional regulator and components of an ABC-like transporter system, followed by two of the latter putative GH130 genes (CALPO_RS0109450 and CALPO_RS0109445) in tandem. These two genes are hereon referred to as Man130C and Man130D. The fifth gene, CALPO_RS0110910 or Man130E, is located elsewhere in the genome and is not situated near ABC transporters.

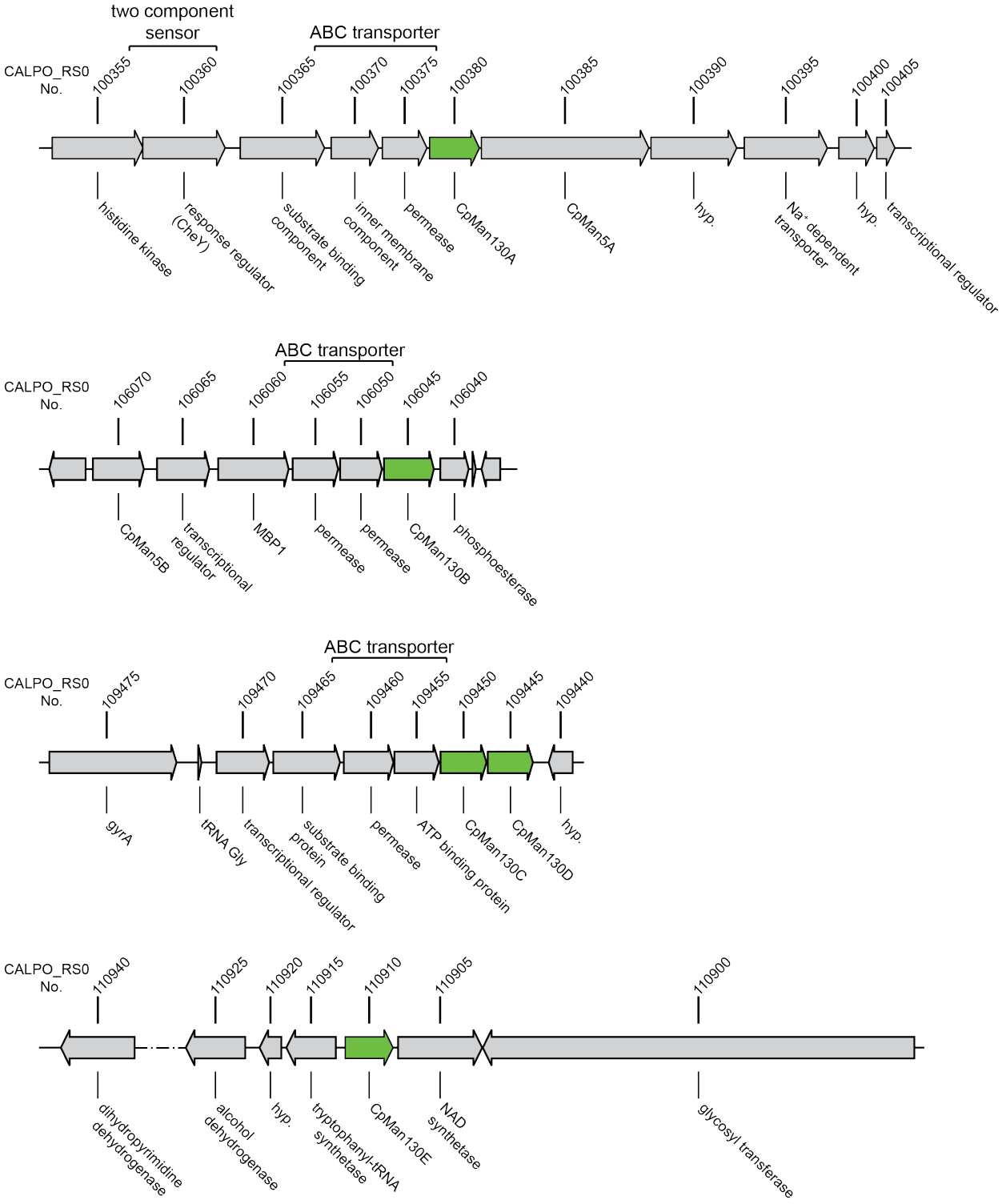


Fig. 4.1 – Five GH130 encoding genes in *C. polysaccharolyticus*
Genomic context for the GH130-family enzyme encoding genes in *C. polysaccharolyticus*, shown in green and annotated according to genome sequence obtained from the Joint Genomic Institute (JGI).

4.2 Biochemical characteristics of the five GH130 proteins

All five GH130 polypeptides encoded in *C. polysaccharolyticus* are composed of a single GH130 module and none possesses a signal peptide, suggesting that each of these enzymes functions in the cytoplasm (Fig. 4.2 A). This characteristic is consistent with other previously characterized GH130-family proteins [7, 9, 13, 67]. Amino acid sequence identities shared by the 5 polypeptides were determined, in addition to comparing each polypeptide with previously characterized GH130 enzymes of both subfamilies GH130_1 and GH130_2 [7, 10, 11, 13, 66] (Table 4.1). The *C. polysaccharolyticus* Man130A and Man130B, located within the Man5A and Man5B gene clusters respectively, share 99% amino acid sequence identity. The genes encoding Man130A and Man130B were previously shown to be up-regulated when *C. polysaccharolyticus* is grown on different mannans as carbon sources [8]. Located *in tandem* in a separate gene cluster, Man130C and Man130D share 61% amino acid sequence identity. When Man130C and Man130D are compared to Man130B, the amino acid sequence identities decline to around 30% and 34%, respectively. Finally, Man130E shares less than 40% of its sequence with any of the other GH130 proteins identified in this organism (Table 4.1).

Each protein was expressed with an N-terminal His_6-tag to facilitate purification by immobilized metal affinity chromatography (IMAC) and a step of heat denaturation was included to denature the mesophilic *E. coli* proteins. Sodium-dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was used to assess the purity of the protein and also to confirm their molecular masses based on calculations from their primary amino acid sequences (Fig. 4.2 B). Each of the proteins was analyzed for their

Table 4.1 – Amino acid sequence identity of the GH130 proteins in *C. polysaccharolyticus* (%)

	CpMan130A	CpMan130B	CpMan130C	CpMan130D	CpMan130E
CpMan130A	**100.00**	99.39	33.33	30.85	36.89
CpMan130B	99.39	**100.00**	33.33	30.85	37.13
CpMan130C	33.33	33.33	**100.00**	61.82	34.75
CpMan130D	30.85	30.85	61.82	**100.00**	37.23
CpMan130E	36.89	37.13	34.75	37.23	**100.00**

Table 4.2 – A comparison of the native molecular weights of the GH130 proteins from *C. polysaccharolyticus*

	Theoretical M.W. (kDa)	Observed on AGF (kDa)	Number of subunits
CpMan130A	37.5	323.81	8.63
CpMan130B	37.3	270.59	7.25
CpMan130C	34.4	77.33	2.25
CpMan130D	34.2	43.46	1.27
CpMan130E	36.0	47.42 or 295.26	1.32 or 8.20

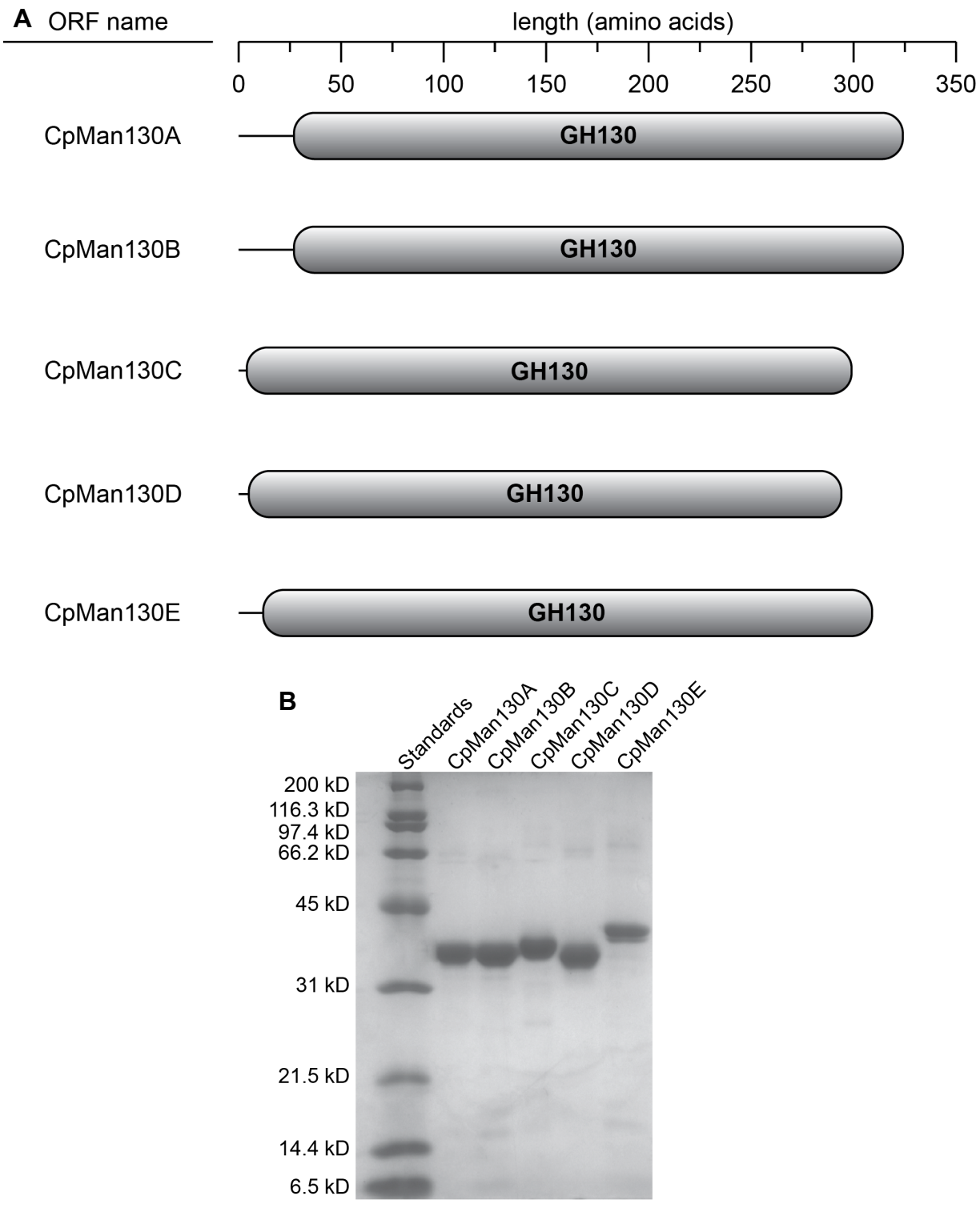


Fig. 4.2 – GH130 enzymes from *C. polysaccharolyticus*
(A) Schematics of the predicted domain organization of the five GH130 proteins (CpMan130A-E) from *C. polysaccharolyticus*. (B) SDS-PAGE analysis of the five GH130 proteins after metal affinity purification.

native molecular mass by analytical gel filtration and their determined weights were compared to the theoretical values in Table 4.2. Chromatograms and other more detailed information can be found among the figures presented in Appendix C. Biochemical characterization of the enzymes was carried out without the excision of the His_6-tag.

4.3 Optimal pH and temperature of the five GH130 enzymes

The pH and temperature optimization was carried out for the five GH130 enzymes by assaying substrate depletion. Mannobiose was utilized as the substrate and the reactions were carried out in presence of inorganic phosphate. Three of the enzymes, Man130C, Man130D and Man130E, did not exhibit cleavage activity with mannobiose, mannotriose or any manno-oligosaccharides with higher degrees of polymerization. Therefore, their optimal pH and temperature values were not determined with this method. Man130A and Man130B exhibited optimal conditions that had little deviation from the optimal growth conditions of *C. polysaccharolyticus*, as can be seen in Table 4.3. Although optimal pH for both enzymes ranged between 6.0 and 6.5, they retained a large percentage of their activity in a wider pH range (Fig. 4.3).

Both Man130A and Man130B exhibited activity in a broad temperature range, retaining over 80% of their activities in temperatures ranging from 50 ºC to around 70 ºC (Fig. 4.3). Although the two enzymes have almost identical amino acid sequences, the optimal temperature of Man130B was 5 °C higher than that of Man130A.

Table 4.3 – pH and temperature optima of GH130 enzymes

Enzyme	Optimum pH	Optimum temperature (ºC)
CpMan130A	6.5	60
CpMan130B	6.0	65
CpMan130C	N.O.*	N.O.*
CpMan130D	N.O.*	N.O.*
CpMan130E	N.O.*	N.O.*

*data could not be collected in this experiment for the enzymes Man130C, Man130D and Man130E because these enzymes exhibited no activity on the commercially available β-1,4-linked manno-oligosaccharides, cello-oligosaccharides, arabino-oligosaccharides or xylo-oligosaccharides.

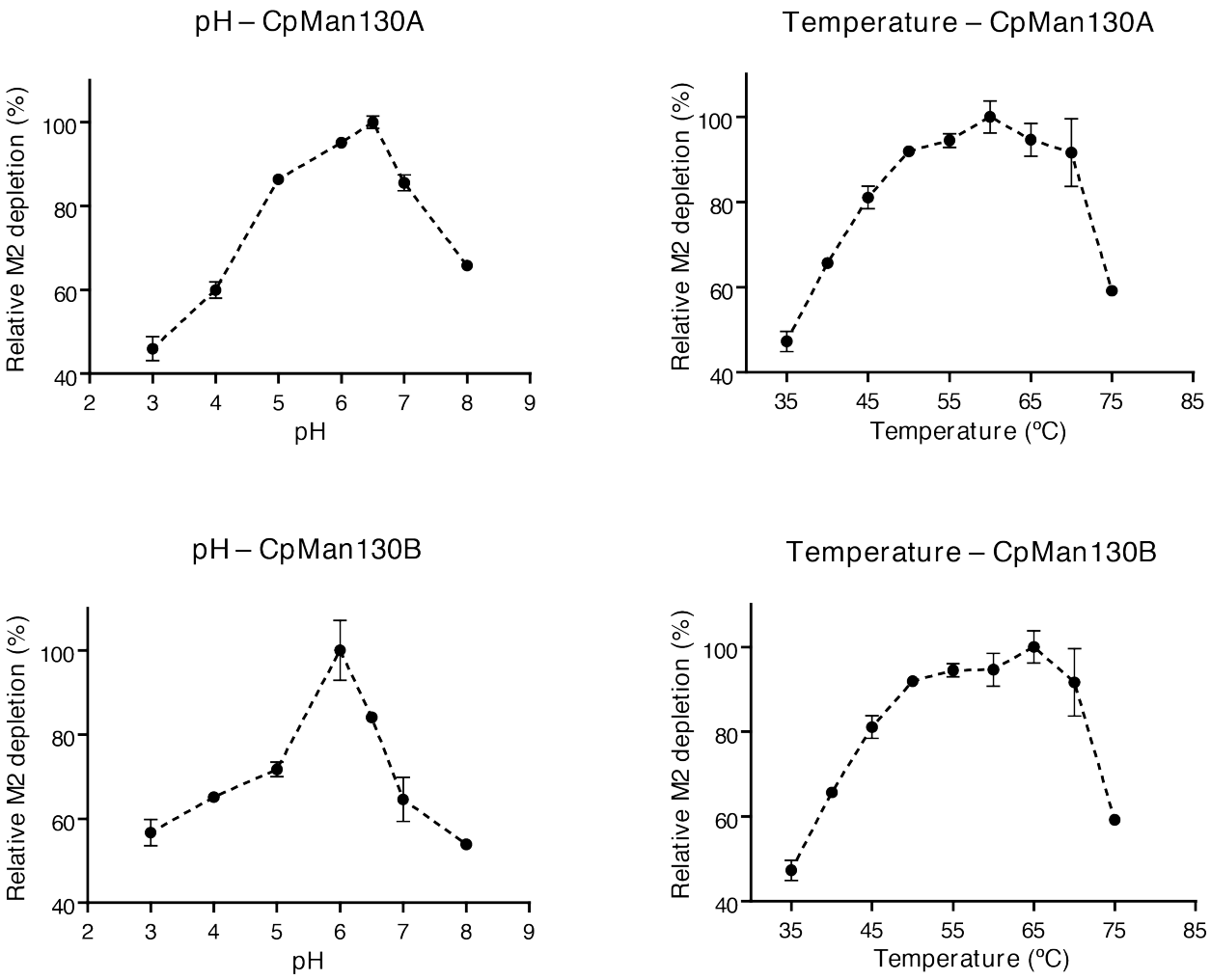


Fig. 4.3 – Determination of optimal pH and temperature for Man130A and Man130B
Optimal pH and temperature conditions for the Man130A and Man130B of *C. polysaccharolyticus*, relative to the highest value. Graphs in the left display the pH and graphs in the right display temperature optimization. Error bars show the S.E.M. of a triplicate experiment.

4.4 Activity of GH130 enzymes in various oligosaccharides

The product release patterns of the five GH130 family enzymes from *C. polysaccharolyticus* were analyzed after incubation at the optimal pH and temperature of each enzyme with different manno-oligosaccharides (DP of from 2 to 6). Man130A and Man130B displayed a similar pattern of product release, which was expected as the two enzymes share an almost identical amino acid sequence. After one hour, most of the oligosaccharides composed of three or more monomeric units were depleted, with mannose and mannobiose being released as end products. When mannobiose was utilized as the substrate, a resulting mixture of mannose and mannotriose was observed (Fig. 4.4A, first two columns). Although the concentrations were too low for detection by HPLC, ^{1}H NMR analysis revealed that α-mannose-1-phosphate is released in approximately equimolar amounts when compared to mannose (Fig. 4.4C). These results are consistent with the activity described for the GH130_2 subfamily, to which both enzymes belong.

Man130C and Man130D did not display detectable activity in any of the tested β-1,4-linked substrates. After 12h of incubation at 65 ºC, trace amounts products with DP of one less than the substrate could be observed for oligosaccharide substrates with a DP of 4 or larger (Fig. 4.4A, last two columns). These data are evidence that β-1,4-linked oligomannans are not the optimal substrate for either of the enzymes. Phylogenetically, both enzymes are not closely related to the subfamilies GH130_2 or GH130_1, which exhibit activity on β-1,4-linked manno-oligosaccharides and on β-1,4-mannosyl-glucose, respectively. In fact, the two polypeptides are more closely related to

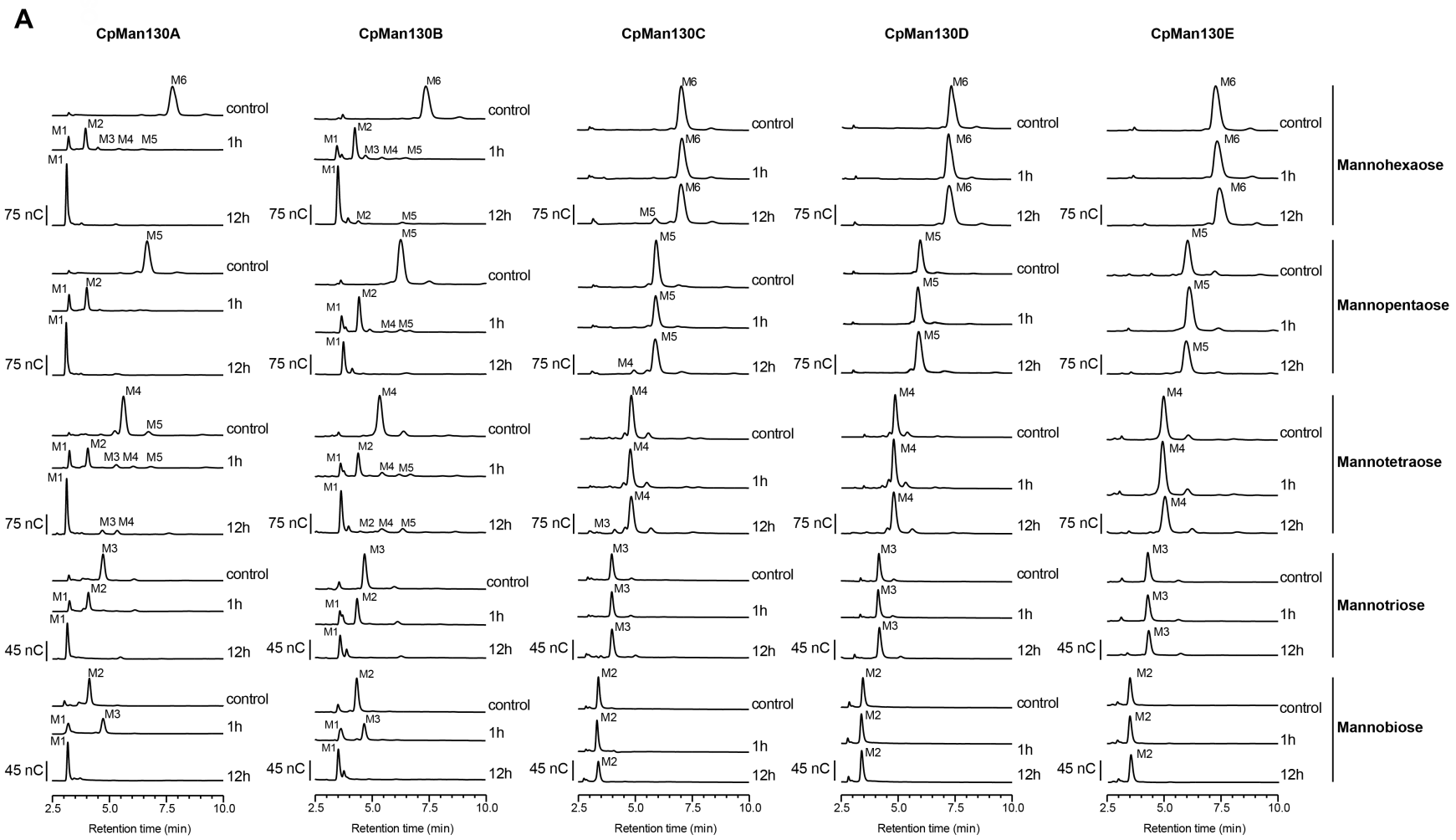
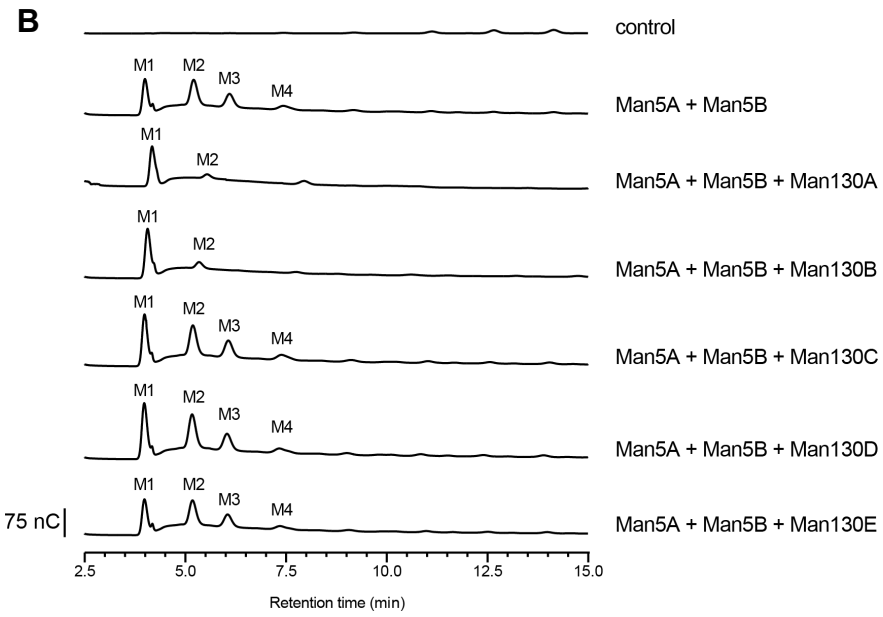


Fig. 4.4 – Product release pattern of the GH130 on MOS

(A) Chromatograms showing product release pattern of CpMan130A, CpMan130B, CpMan130C, CpMan130D and CpMan130E after incubation with β-1,4-manno-oligosaccharides of varying lengths, analyzed by HPAEC-PAD. (B) Chromatograms showing product release pattern of CpMan130A, CpMan130B, CpMan130C, CpMan130D and CpMan130E after incubation with Man5A and Man5B. (C) ^{1}H NMR spectra showing product release patterns of CpMan130A and CpMan130B after incubation with β-1,4-mannobiose. (D) ^{1}H NMR spectra of the synthetic reactions of CpMan130A, CpMan130B, CpMan130C, CpMan130D and CpMan130E after incubation with mannose and α-mannose-1-phosphate for 12h at 65ºC.



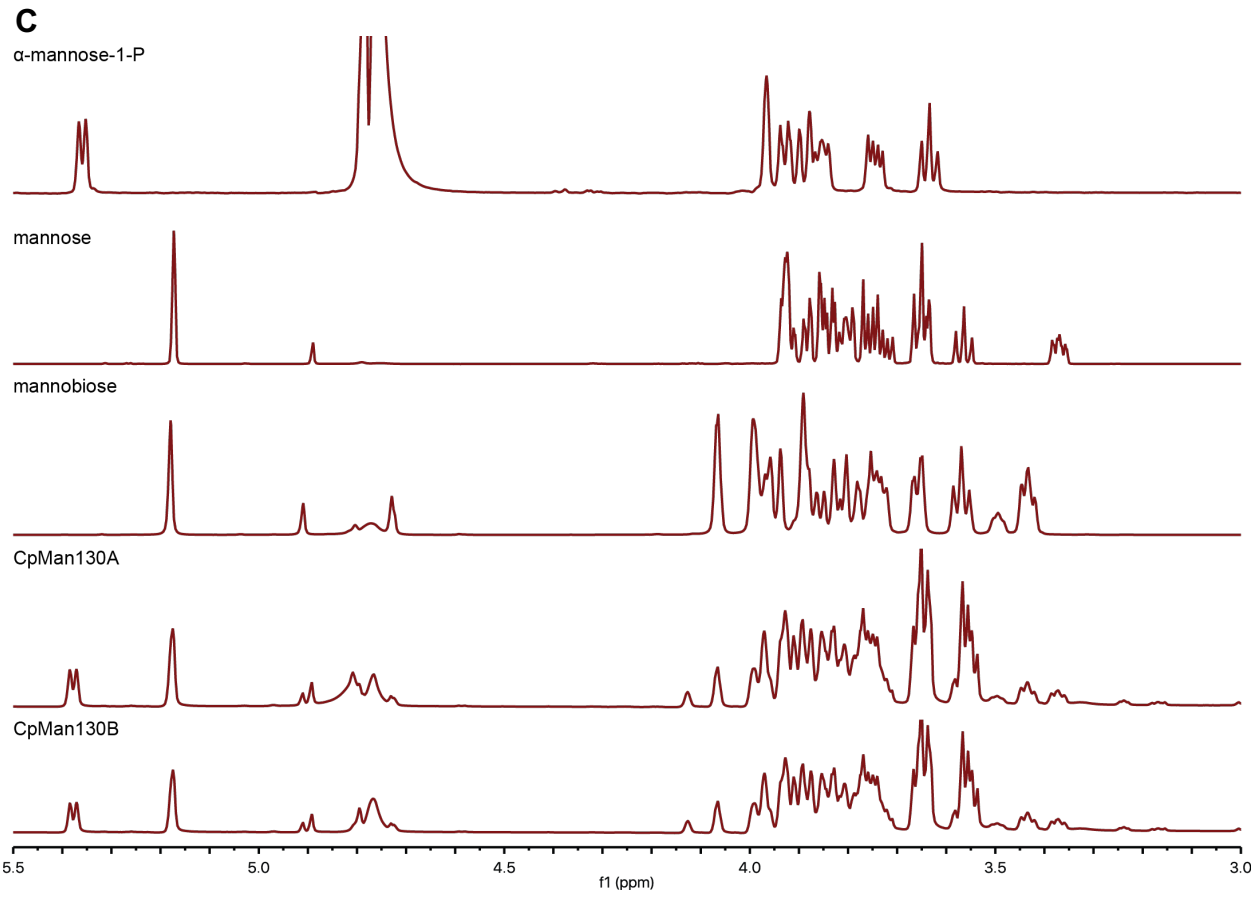


Fig. 4.4 – continued.

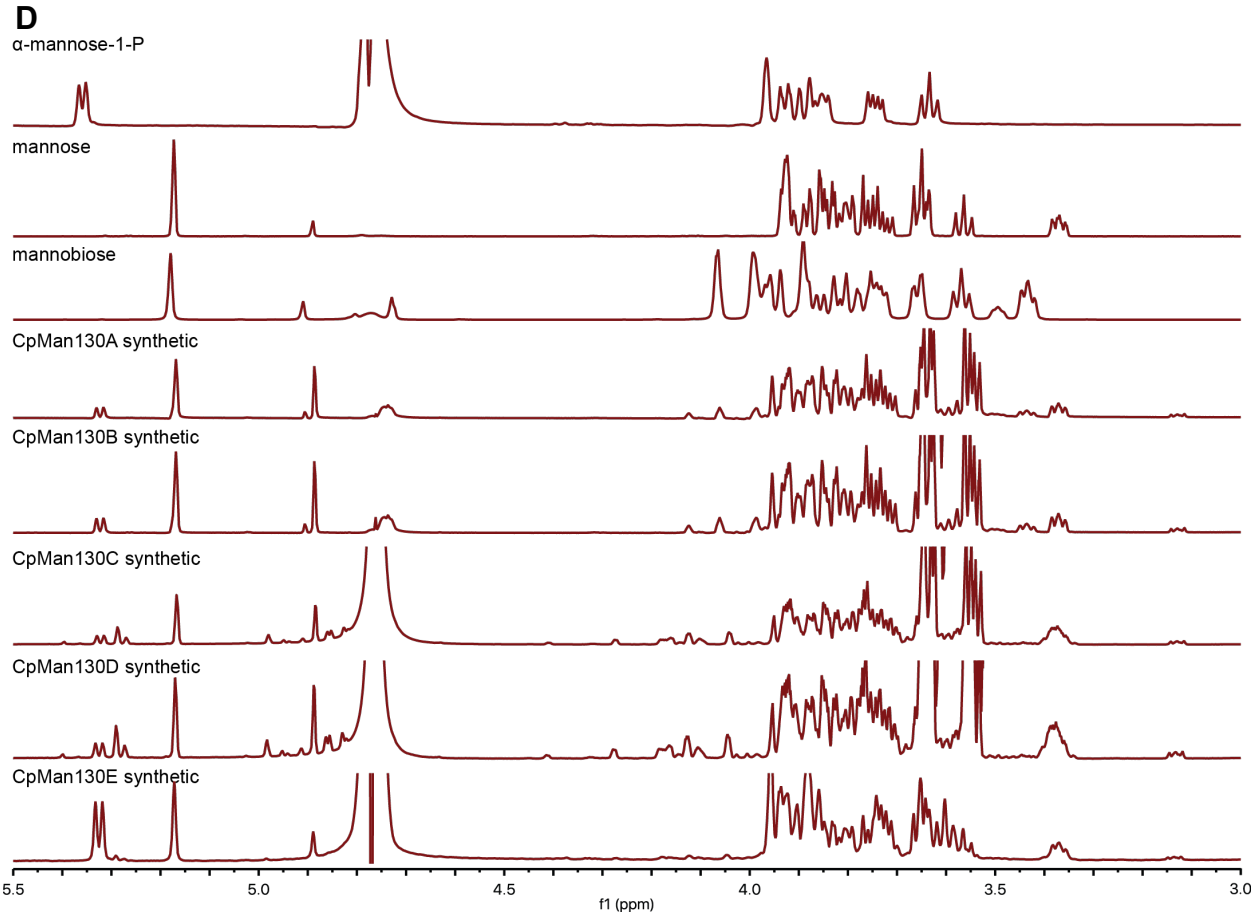


Fig. 4.4 – continued.

two characterized enzymes from *Thermoanaerobacter* sp. X514 [12], and both strain X514 enzymes rather exhibited activity on β-1,2-oligomannans. Accordingly, the formation of trace amounts of a manno-oligosaccharide with a DP of two could be observed when CpMan130C and CpMan130D were incubated with mannose and α-mannose-1-phosphate. Heteronuclear single quantum coherence (HSQC) spectroscopy analysis revealed that the linkages in the produced disaccharides generated different signals from those of β-1,4-mannobiose. Analysis of the NMR shifts suggests that these enzymes have produced a β-1,2-linked disaccharide (Appendix D), as supported by the previously reported chemical shifts for closely related enzymes [12], and this is likely the optimal substrate for the enzyme *in vivo*.

The Man130E of *C. polysaccharolyticus* did not exhibit detectable activity on any of the tested β-1,4-manno-oligosaccharides (Fig. 4.4A), or on any of the tested cello-oligosaccharides, xylo-oligosaccharides or arabino-oligosaccharides with DP ranging from two to six (data not shown). Incubation with mannose and α-mannose-1-phosphate revealed the formation of products with a ^{1}H NMR spectrum that is likely from β-1,2-manno-oligosaccharides, i.e, products very similar to those produced by CpMan130C and Man130D (Fig. 4.4D). CpMan130E produced an NMR signal that was much weaker than those of Man130C and Man130D, which might indicate that it might prefer longer chains that are currently unavailable commercially.

Measurements of the activity in conjunction with the endomannanase Man5A and the β-mannosidase Man5B were performed for each of the five enzymes. Supplementing the endomannanase with either Man130A or Man130B resulted in the degradation of a larger portion of the substrate to mannose and thus a similar

observation to incubating Man5A and Man5B on β-1,4-mannan. By having Man5A, Man5B and either Man130A or Man130B together, it is possible to achieve a more efficient conversion to monomeric sugars after 12h (Fig. 4.4B), besides the generation of the phosphorylated sugar. Man130C, Man130D and Man130E did not exert any measurable effects on the reactions in which they were included, suggesting that they are likely involved in a different degradation pathway. This is in agreement with the suggestion that these enzymes are active on β-1,2-linked oligosaccharides, based on their synthetic reactions, rather than the β-1,4-linked substrates which Man5A and Man5B cleave.

4.5 Isothermal Titration Calorimetry (ITC) of CpMan130B

As previously reported, several GH130 family proteins require the presence of inorganic phosphate in the reaction mixture to exhibit catalytic activity. Hence, ITC was used to investigate the preferred order, if any, in which the phosphate molecule and the carbohydrate substrate bind to the active site of a representative GH130_2 protein. The CpMan130B did not show detectable binding to either mannobiose (Fig. 4.5A) or cellobiose (data not shown) in the absence of inorganic phosphate, but was able to bind to phosphate in the absence of carbohydrate substrates (Fig. 4.5B). When exposed to a mixture of inorganic phosphate and either mannobiose or cellobiose, the protein bound with increased affinity in comparison to phosphate alone. The calorimetry measurements with phosphate and mannobiose displayed heat generation that is likely caused by the cleavage of the β-1,4-mannosidic bond within the active site (Fig. 4.5C),

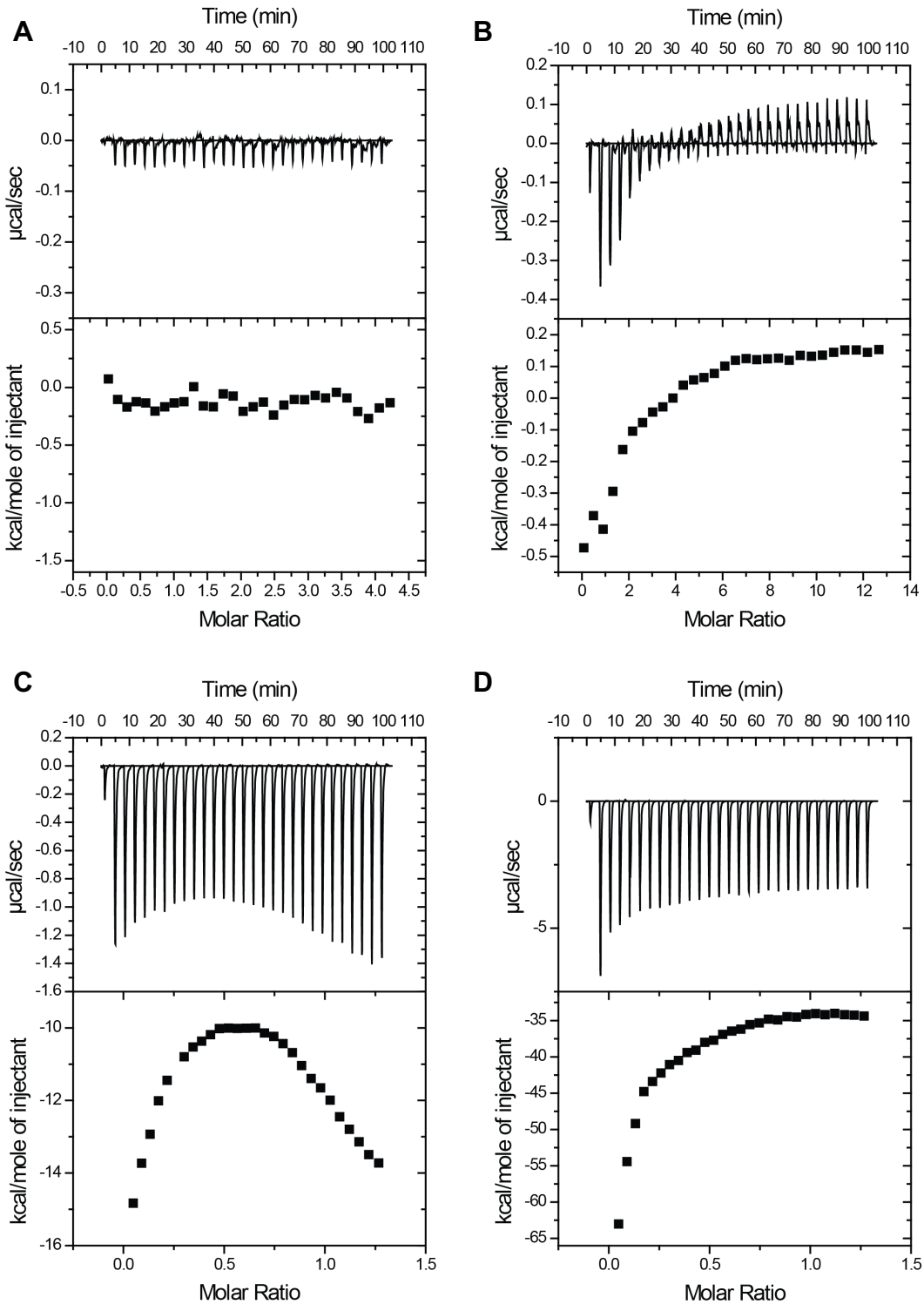


Fig. 4.5 – Substrate binding to CpMan130B
Isothermal titration calorimetry performed with CpMan130B, phosphate and carbohydrates. (A) Incubation with mannobiose in the absence of inorganic phosphate. (B) Incubation with inorganic phosphate in the absence of either cellobiose or mannobiose. (C) Incubation of mannobiose in the presence of inorganic phosphate. (D) Incubation of cellobiose in the presence of inorganic phosphate.

but the same could not be observed when the incubation occurred with cellobiose (Fig. 4.5D). Together, these data support the bi-bi mechanism proposed for the GH130_2 subfamily, as well as the observed lack of activity in the glucose-composed substrates.

4.6 Phylogeny of the GH130 enzymes

To determine whether the studied GH130 polypeptides in *C. polysaccharolyticus* were part of subfamilies GH130_1, GH130_2 or _NC of the GH130 family, a phylogenetic analysis was carried out on the *C. polysaccharolyticus* proteins and the 951 GH130 polypeptides present in the CAZy database. Initially, a tentative phylogeny was carried out with the methodology described by Ladevéze [8] that led to the proposal of the subfamilies, but in this case including all GH130 sequences instead of only those found within the human gut. When the alignment was performed with BLOSUM62 as the predefined distance matrix, the resulting tree led to CpMan130A and CpMan130B in opposing branches (data not shown). This unexpected result led to a further investigation of the methods.

The distance matrix generated by MUSCLE seemed to display a better alignment between these two proteins, which are expected to be clustered very closely together due to their very high sequence identity (Table 4.1). The tree generated using the MUSCLE distance matrix displayed clusters for subfamilies GH130_1 and GH130_2 that matched those obtained by Ladevéze [8] in overall shape (Fig. 4.6) and approximate taxonomic composition, despite the different methodology.

Enzymes in the GH130 family that had their biochemical characterization in the literature are shown as names in the phylogenetic tree. The branch corresponding to

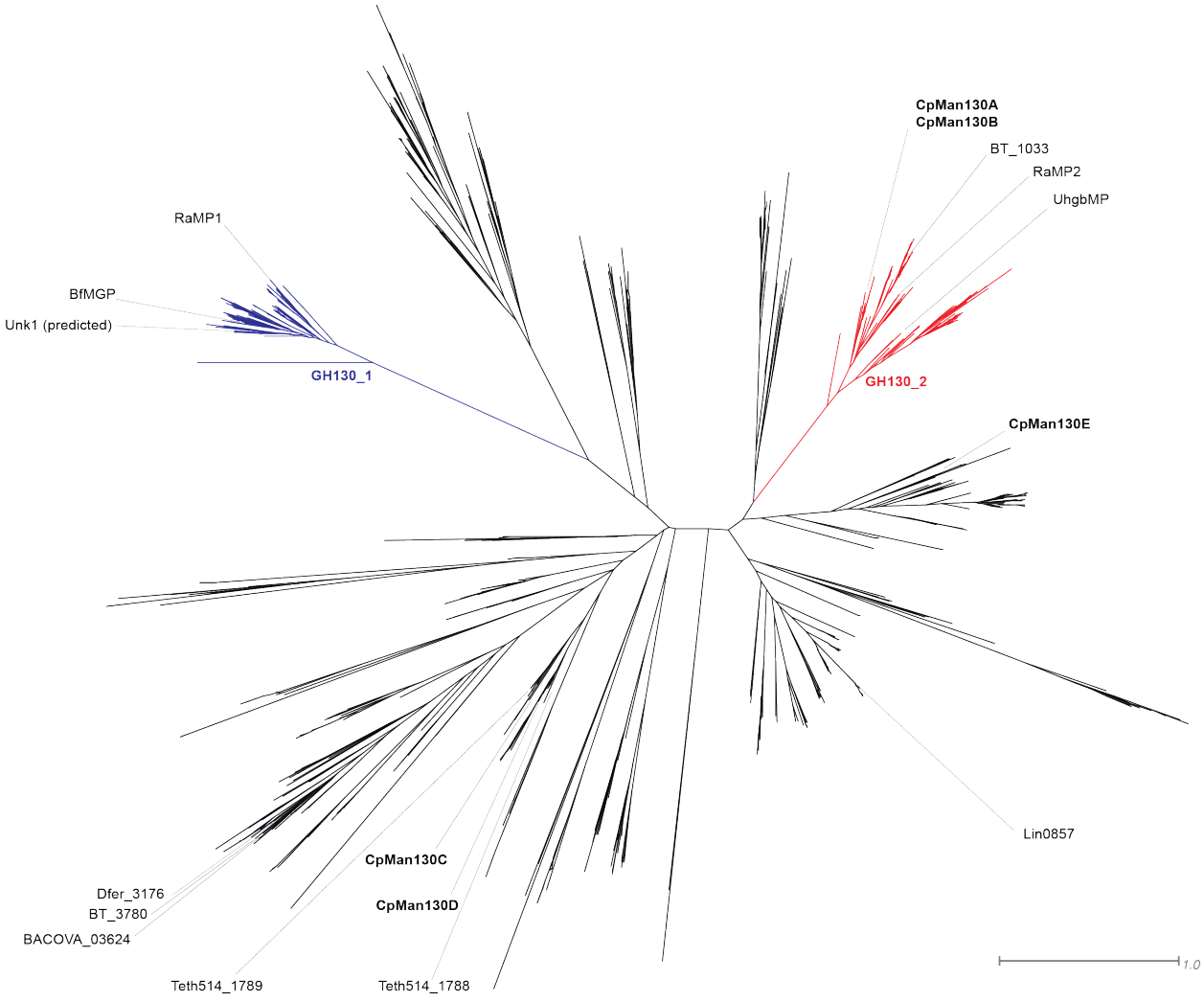


Fig. 4.6 – Phylogeny of the GH130-family proteins
Radial phylogram of the GH130 family containing all characterized enzymes (*http://cazy.org/GH130_characterized.html*), displayed in evidence. Branches corresponding to the subfamilies GH130_1 and GH130_2 are colored in blue and red, respectively. Branches in black correspond to proteins that are not classified into any subfamily (GH130_NC). Enzymes in this study are shown in bold.

the subfamily GH130_1 is colored in blue and evidences its members RaMP1 [10, 66] and BfMGP [8], as well as the putative Unk1 [8]. The branch corresponding to subfamily GH130_2 is colored in red and evidences its members CpMan130A and CpMan130B, along with BT1033 [8], RaMP2 [10, 66] and UhgbMP [7]. Branches in black correspond to proteins that are not classified into any subfamily (GH130_NC) and include the characterized Lin0857 [8], Dfer_3176 [8], Teth514_788 and GH130_1789 [8], BACOVA_03624 and BT3780 [67], along with CpMan130C, CpMan130D and CpMan130E.

Two of the proteins from *C. polysaccharolyticus*, CpMan130A and CpMan130B, were clustered together within subfamily GH130_2, along with UhgbMP [7], BT1033 [11] and RaMP2 [10, 66]. The taxonomic diversity of this subfamily includes enzymes from the thermophilic genera *Spirochaeta*, *Caldicellulosiruptor*, *Mesotoga*, *Fervidobacterium*, *Dictyoglomus*, *Pseudothermotoga* and *Mahella*, as well as from the free-living *Paenibacillus, Spirochaeta* and *Elizabethkingia* and the gut commensals *Bacteroides*, *Ruminococcus*, *Roseburia*, *Clostridium* and *Barnesiella.*

Likewise, CpMan130C and CpMan130D clustered together in close proximity, in one of the branches not assigned to either subfamily (_NC), along two previously characterized proteins, Teth514_1788 and Teth514_1789 from *Thermoanaerobacter* sp. X-514. The two proteins are likely active against β-1,2-linked oligomannans [8], as suggested by the formation of these oligosaccharides when in presence of mannose and α-mannose-1-phosphate. The enzymes in this branch come from thermophilic organisms, including the bacterial and archaeal genera *Thermoanaerobacter*,

Thermoanaerobacterium, *Caldanaerobacter*, *Symbiobacterium*, *Caldicellulosiruptor*, *Kyrpidia* and *Methanothermus*.

The remaining GH130 enzyme from *C. polysaccharolyticus*, i.e., CpMan130E, is located in a gene branch containing no other characterized enzymes. Moreover, a large portion of the branch is populated by polypeptides from thermophiles such as *Rodothermus*, *Thermoanaerobacter* and *Melioribacter*, and methanogenic Archaea such as *Methanosarcina*, and *Methanolobus*. Some enzymes from different *Clostridia* spp. also clustered as close relatives.

It is also worth mentioning that, of the five genes encoding GH130-family proteins within the genome of *C. polysaccharolyticus*, none codes for a protein belonging to the GH130_1 subfamily.

4.7 Structural and computational analyses of Man130A and Man130B

Structural analysis was carried out with the computer generated models. Both Man130A and Man130B displayed the conserved 5-fold β-propeller that is attributed to the enzymes in the GH130 family. The model was generated to include a phosphate molecule and a β-1,4-mannobiose in the active site (Fig. 4.7A). Based on the distance and positioning, as well as sequence conservation, the amino acids likely implicated in the phosphate binding have been identified as R171, R189 and K229 (Fig. 4.7C), and H195 and Y259 have been proposed to play a role in stabilization of the phosphate. Whilst all the listed amino acids are found in the main barrel, the histidine is located in a flexible loop (Fig. 4.7B), and is likely involved in promoting a conformational change upon binding to phosphate. The amino acids likely involved in the stabilization of the

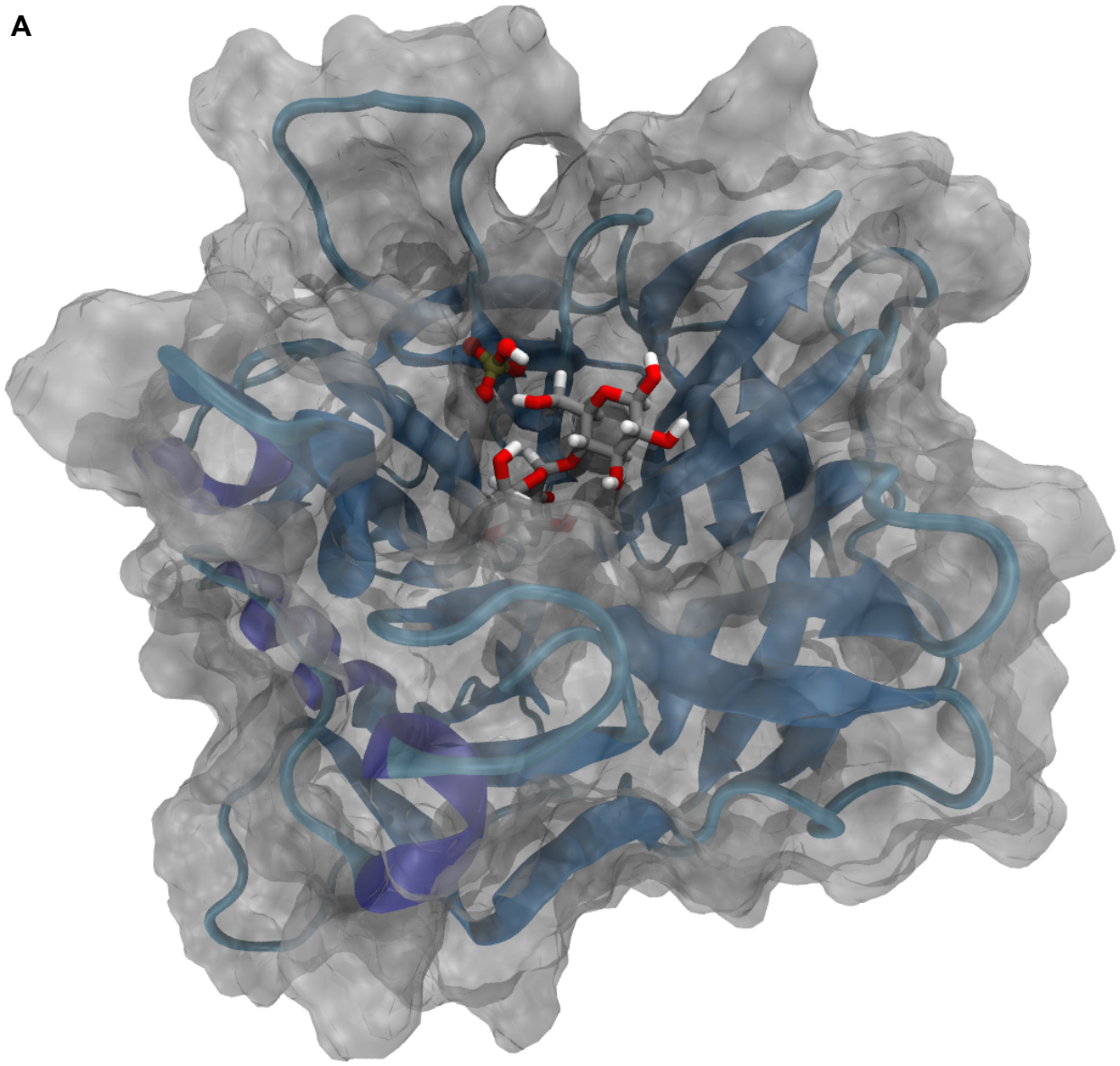


Fig. 4.7 – Computer modeled three-dimensional structure of Man130B
(A) Three-dimensional model of Man130B harboring a phosphate and a β-1,4-mannobiose in the active site. (B) Amino acids involved in the stabilization of the inorganic phosphate are highlighted in the structure. H195, believed to be implicated in conformational changes, is located in an outside loop. (C) Detail of the Hydrogen bonds between the phosphate molecule and its three binding amino acids, R171, R189 and K229.

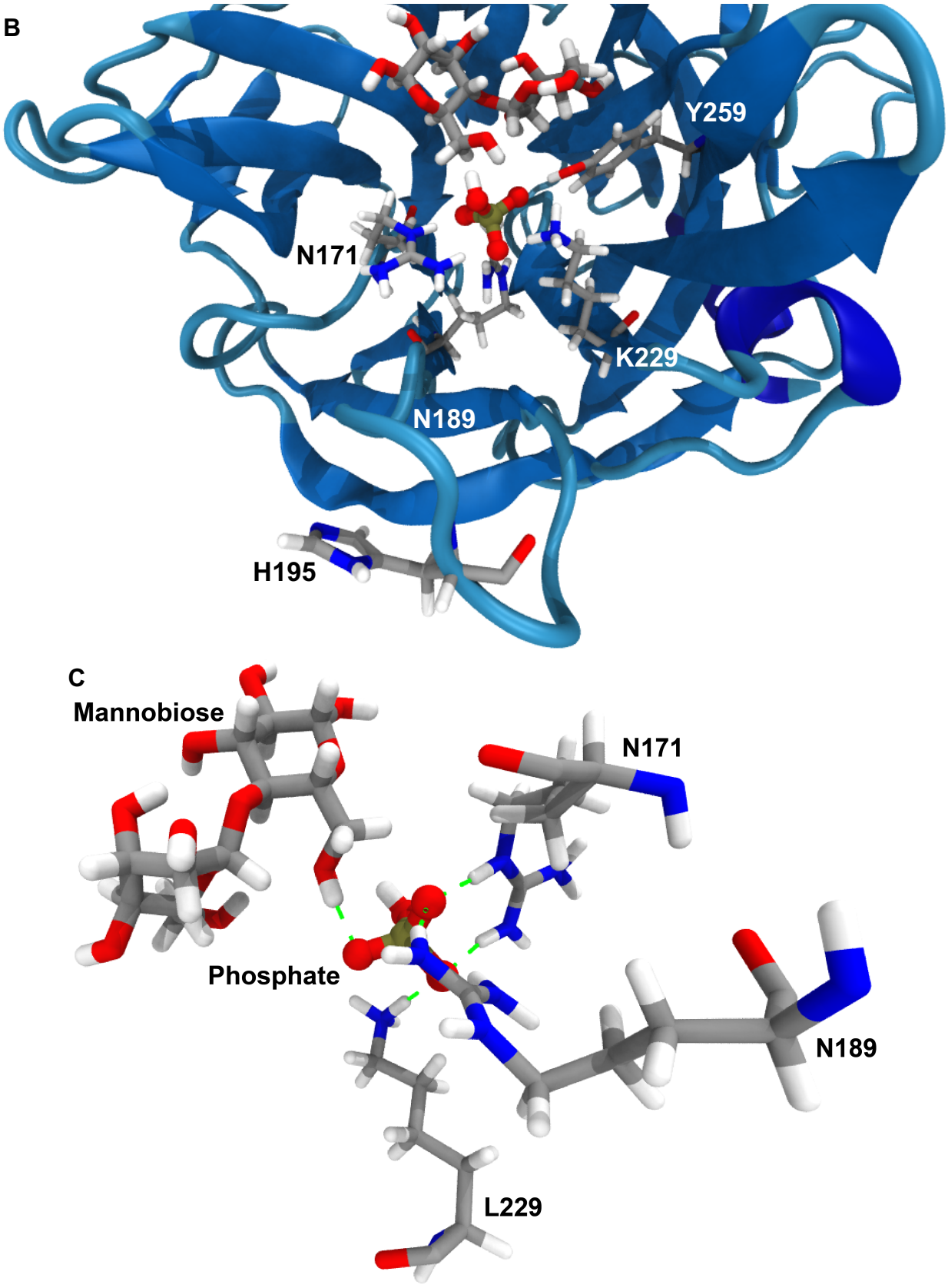


Fig. 4.7 – continued.

Table 4.4 – Amino acids in the active sites of Man130A and Man130B

Enzyme	Phosphate stabilization					Substrate binding			Catalysis
CpMan130A	R171	R189	H195	K229	Y295	N64	D294	D321	D124
CpMan130B	R171	R189	H195	K229	Y295	N64	D294	D321	D124

non-reducing mannoside residue were found to be N64, D294 and D321 of either Man130A or Man130B (Fig. 4.7D). An aspartate located opposite to the phosphate molecule, D124, is believed to be the proton donor. It is likely that the distance between this amino acid and the substrate is reduced when in the active conformation, caused by the binding of phosphate, which could explain the inability to bind to the sugar substrates in the absence of the cofactor. The positions and likely roles of the amino acids hypothesized to be involved in the activity are listed on table 4.4.

CHAPTER 5: DISCUSSION

Thermophilic microorganisms thrive in physically or geochemically extreme conditions that are detrimental to most life on Earth. In order to do so, these organisms have evolved special characteristics that allow growth in nutrient deprived or otherwise metabolically challenging conditions, including the ability to degrade complex substrates and enzymes that are stable under high temperatures. *Caldanaerobius polysaccharolyticus* is an example of a thermophilic bacterium that is able to grow at temperatures above 60 ºC and obtain its nutrients from the degradation of vegetal biomass. This combination of factors makes this organism very attractive for potential industrial application of its cellulolytic and/or hemicellulolytic enzymes.

Of special interest for the biofuels and paper industries, this organism possesses a suite of enzymes that allow the complete degradation of mannans, a major component of the biomass obtained from the reforested softwoods. The presence of a bifunctional extracellular endomannanase (CpMan5A) and an intracellular bifunctional β-mannosidase (CpMan5B) allow this organism to degrade the majority of the long chain mannans into shorter chain oligosaccharides. The specificity of the Man5B for oligosaccharides has been reported to be lower for chains with degrees of polymerization (DP) of two or three when compared to chains with larger DPs. An alternative degradation pathway for this bacterium to degrade low DP oligosaccharides into monomeric units is by utilizing complementary proteins, such as mannoside phosphorylases from the GH130 family.

5.1 The GH130 genes of *C. polysaccharolyticus* ATCC BAA-17

The genome of *C. polysaccharolyticus* contains genes that encode five proteins of the glycoside hydrolase family 130 (GH130), two of which are located in the same gene clusters as the genes for Man5A and Man5B. This genomic arrangement might indicate a functional connection between these enzymes in the metabolism of *C. polysaccharolyticus*. The GH130 family is divided into two main subfamilies, GH130_1 and GH130_2, with different substrate specificities, and a third and larger group of unclassified enzymes or GH130_NC [7]. The subfamily GH130_1 has been shown to act on the β-1,4-linkages between mannoside and glucoside residues, and has been proposed to release D-glucose and α-mannose-1-phosphate [13]. Likewise, subfamily GH130_2 has been proposed to have a similar mechanism; however, they only cleaved the linkages in β-1,4-linked mannoside residues, such as those of commonly found in manno-oligosaccharides from degradation of mannans [7]. The third and last group, containing the majority of the proteins classified as GH130, is composed of putative enzymes of varying substrate specificity and reaction mechanisms [7, 12, 67, 69]. Members from the GH130_NC originate from a plethora of microorganisms, most of which thrive under anaerobic conditions.

The presence of five copies of genes encoding GH130 family polypeptides within the genome of *C. polysaccharolyticus* suggests that they play different metabolic roles. This hypothesis is reinforced by an overall amino acid sequence identity lower than 40%, with the exception of CpMan130A and CpMan130B, which share amino acid sequence identity of 99.4%.

A phylogenetic tree containing all GH130 sequences available at the CAZy database was assembled and CpMan130A and Man130B, both being part of subfamily GH130_2, were located in close proximity. The *C polysaccharolyticus* Man130C, Man130D and Man130E, in contrast, clustered with the larger _NC group. CpMan130C and Man130D were located in close proximity and were situated near a pair of characterized enzymes from *Thermoanaerobacter* sp. X-514, with activity on β-1,2-linked manno-oligosaccharides [12]. It is of note that CpMan130E was located in isolation of any previously characterized enzymes.

5.2 CpMan130A and CpMan130B are part of the GH130_2 subfamily

The two almost identical CpMan130A and CpMan130B were only slightly different at the C-terminus. Both enzymes exhibited activity on β-1,4-linked manno-oligosaccharides of lengths ranging from two to six monomeric units, which is in agreement with the proposed activity for subfamily GH130_2. Both enzymes require inorganic phosphate for catalytic activity, and the proteins are unable to bind the substrate in the absence of inorganic phosphate (Fig. 4.5A). Moreover, as phosphate alone is able to bind to the active site of the protein (Fig. 4.5B), it is likely that, once the phosphate is stably positioned within the binding pocket (Y259, R171, R189 and K229), the histidine situated in the 190-200 loop (H195) is shifted to the proximity of phosphate, causing the protein to undergo a conformational change. The conformational change then allows N064, D294 and D321 to interact with the mannoside residue in the -1 sub-site.

When the enzymes were incubated with mannobiose for one hour in the presence of phosphate, the resulting products quantifiable by HPAEC were mannose and mannotriose, whereas the incubation for the same time with longer oligosaccharides resulted in a mixture of mannobiose and mannose. Ultimately, the majority of the HPAEC-quantifiable products consisted of mannose after longer incubation times. When analyzed in shorter incubation times the product composition displayed an array of oligosaccharides of varying lengths, generally ranging from monomeric units to a DP of one longer than the initial substrate.

Although no quantifiable amount of α-mannose-1-phosphate was detected by HPAEC-PAD analysis, the ^{1}H NMR spectrum for the incubation with mannobiose revealed that α-mannose-1-phosphate is released in approximately equimolar amounts when compared to mannose. This indicates that the phosphorolysis of the non-reducing end of a β-1,4-linked n-oligosaccharide by these proteins proceeds with an overall retention of anomeric configuration, leading to a molecule of α-mannose-1-phosphate and a β-1,4-linked oligosaccharide with a DP of one smaller (n-1) than that of the substrate. The release of the phosphorylated sugar is metabolically beneficial to the organism, as the compound can be isomerized into mannose-6-phosphate and shuttled into the central metabolism without the necessity of consuming ATP to phosphorylate the sugar.

5.3 CpMan130C and CpMan130D are members of GH130_3 subgroup

The genome of *C. polysaccharolyticus* harbors a gene cluster containing two *in tandem* genes encoding proteins in the GH130 family, which were identified as part of

the GH130_NC subgroup by phylogenetic analysis. CpMan130C and CpMan130D share less than 62% amino acid sequence identity and exhibited no measurable activity in any of the β-1,4-linked oligosaccharides tested, regardless of their sugar composition or degree of polymerization. A comparison revealed an amino acid sequence identity of over 80% between each of the enzymes from *Caldanaerobius* and their counterparts from *Thermoanaerobacter*, Teth514_1789 and Teth514_1788, respectively.

The enzymes from *Thermoanaerobacter* exhibited substrate specificity towards β-1,2-linked oligomannans with different chain lengths, one of them acting more efficiently on oligosaccharides with a DP of 2 (Teth514_1789) whereas the other acting more efficiently on oligosaccharides with a DP of 3 or greater (Teth514_1788). Both enzymes have been implicated in the biosynthesis of GDP-D-mannose based on their genomic context. This hypothesis was tested with the two GH130_NC members from *C. polysaccharolyticus*.

The two enzymes, Man130C and Man130D, exhibited activities similar to those observed for their counterparts in the *Thermoanaerobacter* species. Phylogenetic analysis placed the *C. polysaccharolyticus* enzymes in two different, but closely related, branches of the tree (Fig. 5.2A). The composition of these branches revealed that a large number of the organisms that possessed proteins clustering in one of the branches also harbored their counterparts from the other branch, as noted in Fig. 5.2B.

There is a trend among species with enzymes in this group where they possess one copy of each of the two genes encoding the complementary proteins [12]. Furthermore, both copies are usually situated *in tandem* in a partially conserved gene cluster, preceded by a putative ABC transporter, as previously reported for 6 species of

the genus *Thermoanaerobacter* [8] and now for *C. polysaccharolyticus*. This pattern can also be observed in four *Thermoanaerobacterium* species (*T. saccharolyticum, T. xylanolyticum,* and two *T. thermosaccharolyticum* strains), *Symbiobacterium thermophilum*, *Caldanaerobacter subterraneus* and *Caldicellulosiruptor hydrothermalis* (Fig. 5.2C). The last organism also contains a third copy of a similar gene (Appendix E). The enzymes from the six *Thermoanaerobacter* species were implicated in the biosynthesis of GDP-D-mannose [8], nevertheless there is no evidence by genomic context that those from *C. polysaccharolyticus* play a similar role, as a gene encoding the conserved α-mannose-1-phosphate guanylyltransferase is absent from the vicinities of the GH130.

It is proposed here that the shared characteristics allows grouping of these enzymes in a new subgroup, GH130_3, comprised of enzymes from thermophilic microorganisms and that produced β-1,2-manno-oligosaccharides when incubated with mannose and α-mannose-1-phosphate. of varying lengths. As different enzymes are specific for oligosaccharide with different lengths, a pattern has been observed where species harbor more than one gene encoding enzymes within this group, often *in tandem* and preceded by genes for an ABC-like transporter. The consensus sequence for the group, obtained my multiple sequence alignment, can be seen on figure 5.1C.

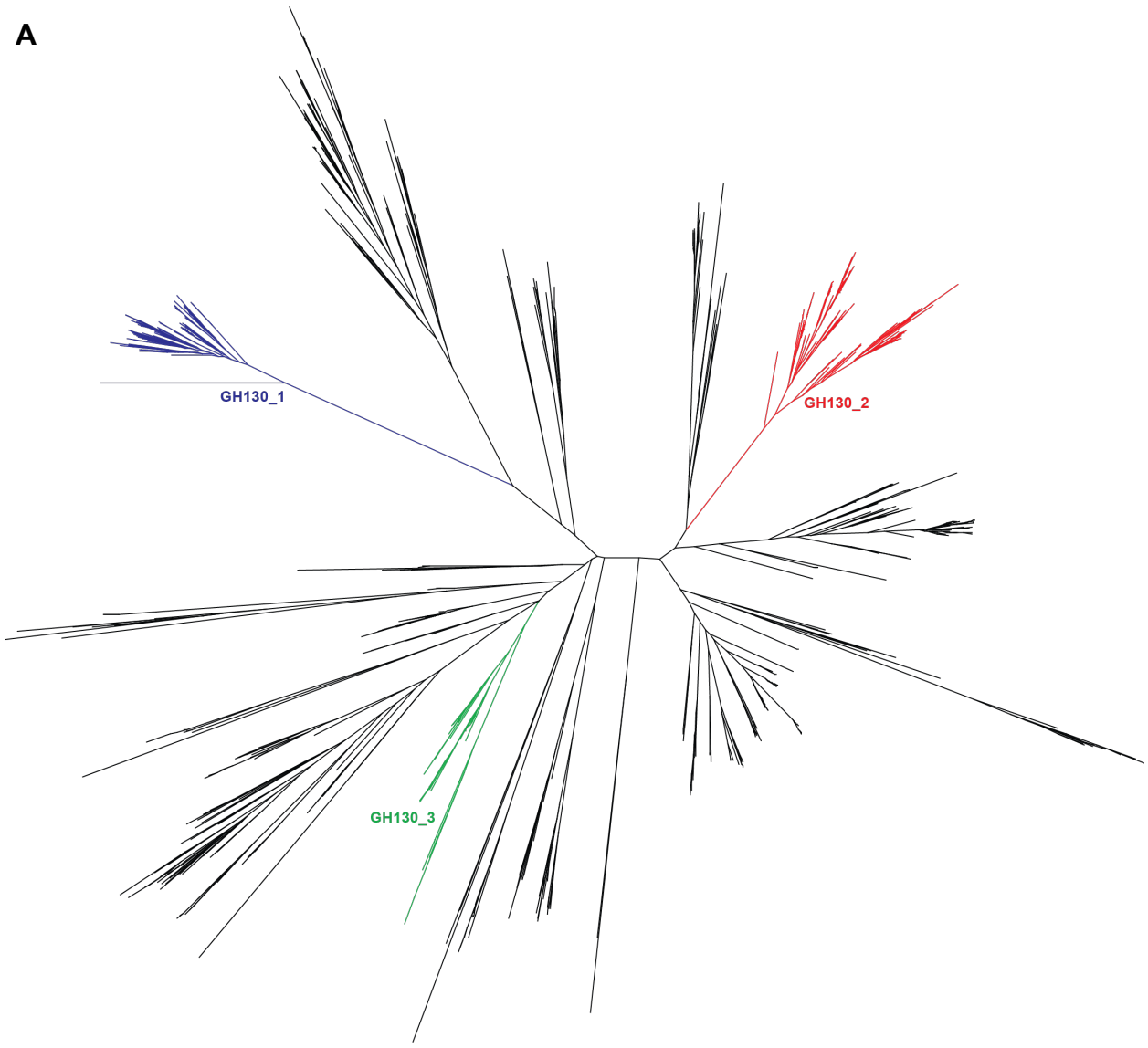
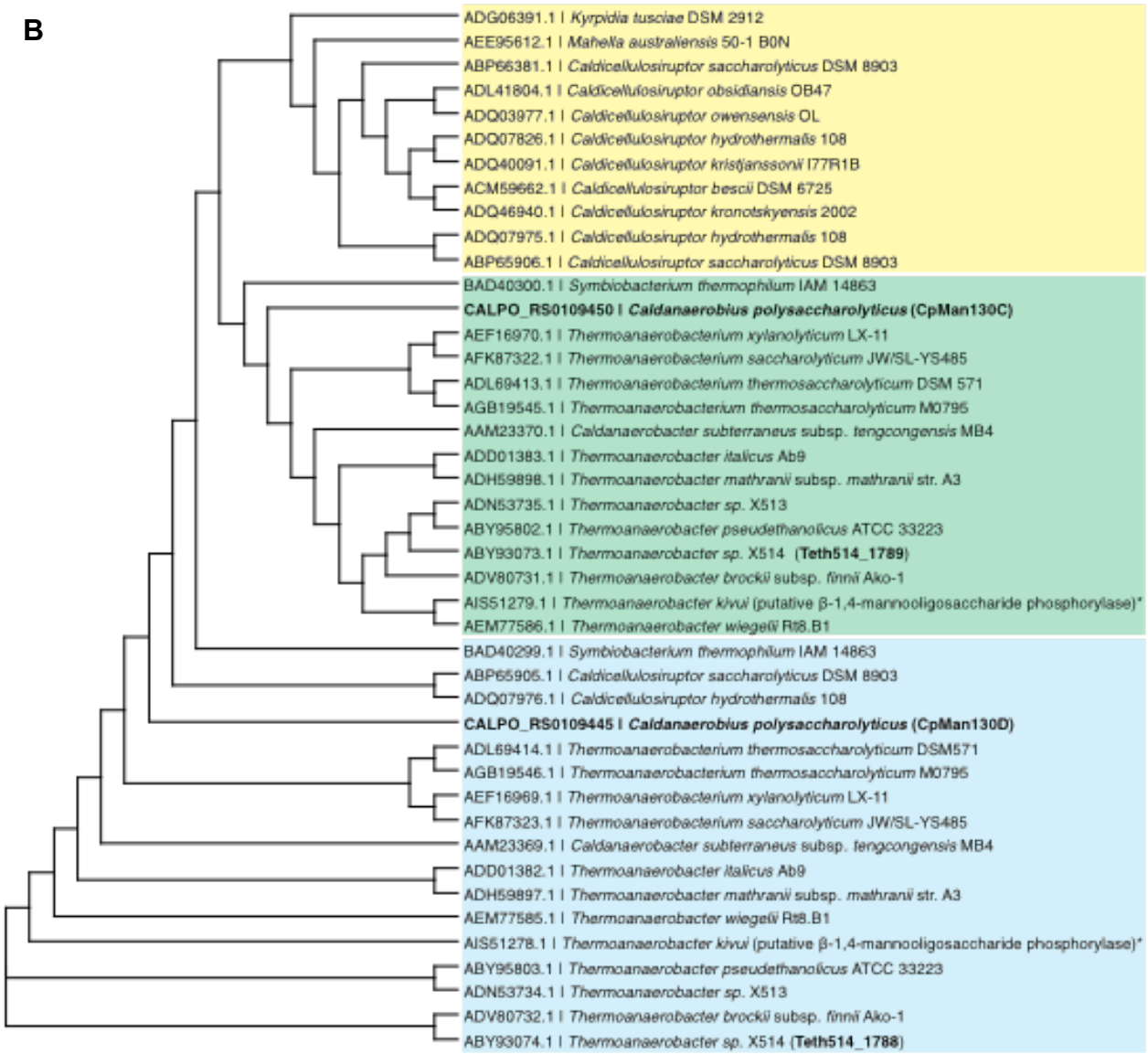


Fig. 5.2 – Phylogeny of the proposed GH130_3 subgroup
(A) Radial phylogram of the GH130 family of enzymes. Branches corresponding to the subfamilies GH130_1 and GH130_2 are colored in blue and red, respectively, and the branch in shown in green represents the proposed GH130_3 subgroup. Branches in black correspond to proteins that are not classified into any subfamily (GH130_NC). (B) Cladogram containing the members of the proposed subgroup. The three colored blocks represent the three branches seen in the phylogenetic tree. Characterized enzymes are shown in bold and enzymes marked with * have been assigned as β-1,4-manno-oligosaccharide phosphorylases by a bioinformatics approach instead of β-1,2-manno-oligosaccharide phosphorylases. (C) Consensus sequence for the proposed GH130_3 subgroup.

B



C

```
>Consensus sequence for subgroup GH130_3
M--L-R----P-L-P-----WZ--AVFN----------J-YR--B--------------------G-A-S-BG--F-R
---P--------QE--G-ED-R----------YT--------B-R-----S--L--W--------DEPNKN--L-----
-G-Y---HTT----W-A-S-D---W--H----------W-----G--GPP-----G----YH--------YRLG--L-D-
-BP-----R----JJEPE--WE--G---NV-FS-G----------YYG--D--JGV----------------
```

Fig. 5.2 – continued.

5.4 CpMan130E is likely related to β-1,2-mannoside metabolism

Located in a cluster with no other genes clearly related to mannan metabolism, CpMan130E did not exhibit measurable activity towards any β-1,4-linked manno-oligosaccharide. The ^{1}H NMR spectra of products from the synthetic reaction of the enzyme revealed very similar pattern to that observed with CpMan130C and Man130D, suggesting that this enzyme is likely involved in the metabolism of β-1,2-linked manno-oligosaccharides. The signals generated by the synthetic reaction were very weak compared to those generated by Man130C and Man130D. There is a possibility that CpMan130E cleaves β-1,2-linked oligosaccharides with a larger degree of polymerization. Hence, the active site architecture would require binding of oligosaccharides and this would decrease the efficiency of the synthetic reaction with monomeric units.

To further test this hypothesis, longer β-1,2-linked oligosaccharides would need to be synthesized, purified as previously described [68] to be utilized as substrate for the synthetic reaction of CpMan130E. Obtaining the three-dimensional structure for this protein might also aid in identifying key features that could point towards specificity for a specific degree of polymerization.

5.5 Overview of the mannan metabolism by *C. polysaccharolyticus*

We propose the following amendment for the pathway of mannan utilization by *C. polysaccharolyticus*. Initial degradation of β-1,4-mannan occurs outside the cell by the endo-mannanase CpMan5A [15, 16]. The generated oligosaccharides enter the cell via an ABC-like transporter and can be degraded either by the β-mannosidase CpMan5B

into shorter manno-oligosaccharides (MOS) and, to a lesser extent, mannose [18], or by the mannosyl-phosphorylases CpMan130A and CpMan130B. When the degradation is carried out by CpMan5B, a hexokinase (HK) utilizes an ATP to phosphorylate one of the released mannose residues to α-mannose-6-phosphate and ADP [11, 18]. CpMan130A and CpMan130B can degrade the MOS directly to mannose and α-mannose-1-phosphate. The mannose follows the same pathway described for CpMan5B, but a phosphomannomutase (PMM) converts the released α-mannose-1-phosphate to α-mannose-6-phosphate [11, 12], preserving ATP (Fig. 5.3). A mannose-6-phosphate isomerase (MPI) then converts the mannose-6-phosphate to fructose-6-phosphate, which proceeds through glycolysis into the central metabolism [11].

Similarly, β-1,2-mannan is initially degraded outside the cell by a β-1,2-mannanase, generating β-1,2-manno-oligosaccharides which are transported into the cell by an ATP-like transporter. Once inside the cell, these oligosaccharides are further degraded into smaller saccharides by CpMan130E and, sequentially, to mannose and α-mannose-1-P by CpMan130C and Man130D. The released mannose is then phosphorylated by a hexokinase leading to α-mannose-6-phosphate. Like previously described for the β-1,4-mannan pathway [11] mannose-1-P is converted to α-mannose-6-P by a phosphomannomutase. A mannose-6-phosphate isomerase (MPI) converts the mannose-6-phosphate to fructose-6-phosphate, which is shuttled through glycolysis into the central metabolism.

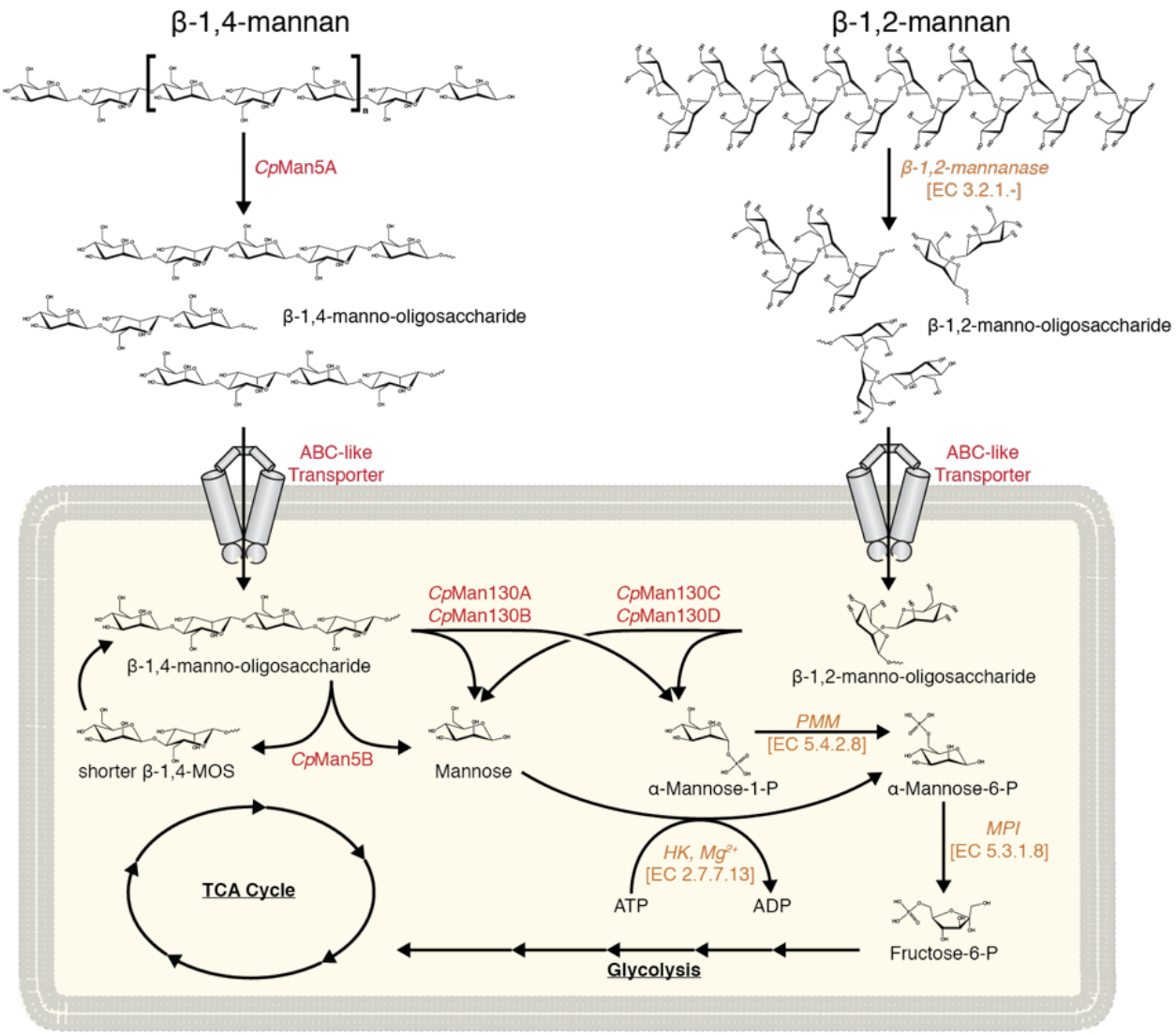


Fig. 5.3 – Proposed metabolism of mannan by *C. polysaccharolyticus* ATCC BAA-17
Schematic metabolism of different types of mannan by *C. polysaccharolyticus* ATCC BAA-17. Degradation of β-1,4-mannan occurs outside the cell by the predicted extracellular endo-mannanase *Cp*Man5A. The generated oligosaccharides enter the cell via ABC-like transporter and can be degraded by the β-mannosidase *Cp*Man5B into shorter manno-oligosaccharides (MOS) and, to a lesser extent, mannose. Then, a hexokinase (HK) converts mannose in an ATP-dependent mechanism into α-mannose-6-P. Man130A and Man130B degrade the MOS directly to mannose and α-mannose-1-P, preserving ATP in the process. Likewise, degradation of β-1,2-mannan is carried out by a β-1,2-mannanase, generating β-1,2-MOS which are transported into the cell by an ATP-like transporter, where they are further degraded into mannose and α-mannose-1-P. The produced mannose-1-P is then converted to α-mannose-6-P by a phosphomannomutase (PMM). A mannose-6-phosphate isomerase (MPI) converts the mannose-6-P to fructose-6-P, which is then shuttled through glycolysis into the central metabolism.

CHAPTER 6: CONCLUSIONS

Caldanaerobius polysaccharolyticus, previously shown to possess the enzymes required to grow on mannan, in the form of an endomannanase (Man5A) and β-mannosidase (Man5B), also harbors five genes encoding GH130 proteins, implicated in the metabolism of different types of manno-oligosaccharides to mannose and α-mannose-1-phosphate. Little is understood about this recently discovered glycoside hydrolase family and we compared the roles of the enzymes identified in this bacterium.

The presence of five different genes encoding proteins in the GH130 family suggests they play different roles in the metabolism of *C. polysaccharolyticus*. Two proteins in the GH130_2 subfamily, Man130A and Man130B, are both located within the gene clusters containing Man5A and Man5B, respectively, and both exhibited substrate specificity towards β-1,4-linked manno-oligosaccharides. Together, these enzymes act to fully degrade β-1,4-mannan to their monomeric sugars and their phosphorylated counterparts.

A cluster containing a conserved gene arrangement in which two copies of GH130 encoding genes, Man130C and Man130D, are preceded by an ABC-like transporter, is present in this organism. Similarly organized clusters have been observed in some other thermophilic bacteria, including the genera *Thermoanaerobacter*, *Caldicellulosiruptor*, *Thermoanaerobacterium* and *Symbiobacter*. Moreover, the amino acid sequence identity between the two enzymes surpassed 60% and the identity of each of these proteins to their counterparts in the aforementioned genera is as high as 80%. These counterparts include the two characterized enzymes

from *Thermoanaerobacter* sp. X-514, which also synthesized β-1,2-linked manno-oligosaccharides when incubated with mannose and α-mannose-1-phosphate. Thus, we proposed that these enzymes could be joined to originate the GH130_3 subgroup.

Although the fifth enzyme, CpMan130E, remains to be biochemically characterized, it is clear this protein is not directly associated with the metabolism of β-1,4-linked manno-oligosaccharides. A better understanding of the function of this enzyme may be achieved based on insights provided by further testing with substrates that are not commercially available. Determining its three-dimensional structure and utilizing a computational approach may also aid in determining the active site architecture to establish which degree of polymerization is preferably utilized by this enzyme.

REFERENCES

1. de Vries, R.P. and J. Visser, *Aspergillus enzymes involved in degradation of plant cell wall polysaccharides.* Microbiol Mol Biol Rev, 2001. **65**(4): p. 497-522.

2. Himmel, M.E., et al., *Biomass recalcitrance: engineering plants and enzymes for biofuels production.* Science, 2007. **315**(5813): p. 804-7.

3. Meyer, A.S., L. Rosgaard, and H.R. Sørensen, *The minimal enzyme cocktail concept for biomass processing.* Journal of Cereal Science, 2009. **50**(3): p. 337-344.

4. Gilbert, H.J., *The Biochemistry and Structural Biology of Plant Cell Wall Deconstruction.* Plant Physiology, 2010. **153**(2): p. 444-455.

5. Lynd, L.R., et al., *Microbial cellulose utilization: fundamentals and biotechnology.* Microbiol Mol Biol Rev, 2002. **66**(3): p. 506-77, table of contents.

6. Moreira, L.R. and E.X. Filho, *An overview of mannan structure and mannan-degrading enzyme systems.* Appl Microbiol Biotechnol, 2008. **79**(2): p. 165-78.

7. Ladeveze, S., et al., *Role of glycoside phosphorylases in mannose foraging by human gut bacteria.* J Biol Chem, 2013. **288**(45): p. 32370-83.

8. Senoura, T., et al., *New microbial mannan catabolic pathway that involves a novel mannosylglucose phosphorylase.* Biochem Biophys Res Commun, 2011. **408**(4): p. 701-6.

9. Kawahara, R., et al., *Metabolic mechanism of mannan in a ruminal bacterium, Ruminococcus albus, involving two mannoside phosphorylases and cellobiose 2-epimerase: discovery of a new carbohydrate phosphorylase, beta-1,4-mannooligosaccharide phosphorylase.* J Biol Chem, 2012. **287**(50): p. 42389-99.

10. Nihira, T., et al., *Discovery of beta-1,4-D-mannosyl-N-acetyl-D-glucosamine phosphorylase involved in the metabolism of N-glycans.* J Biol Chem, 2013. **288**(38): p. 27366-74.

11. Chiku, K., et al., *Discovery of two beta-1,2-mannoside phosphorylases showing different chain-length specificities from Thermoanaerobacter sp. X-514.* PLoS One, 2014. **9**(12): p. e114882.

12. Nakae, S., et al., *Structure of novel enzyme in mannan biodegradation process 4-O-beta-D-mannosyl-D-glucose phosphorylase MGP.* J Mol Biol, 2013. **425**(22): p. 4468-78.

13. Cann, I.K., et al., *Characterization of two novel saccharolytic, anaerobic thermophiles, Thermoanaerobacterium polysaccharolyticum sp. nov. and Thermoanaerobacterium zeae sp. nov., and emendation of the genus Thermoanaerobacterium.* Int J Syst Evol Microbiol, 2001. **51**(Pt 2): p. 293-302.

14. Cann, I.K., et al., *Molecular cloning, sequencing, and expression of a novel multidomain mannanase gene from Thermoanaerobacterium polysaccharolyticum.* J Bacteriol, 1999. **181**(5): p. 1643-51.

15. Han, Y., et al., *Comparative analyses of two thermophilic enzymes exhibiting both beta-1,4 mannosidic and beta-1,4 glucosidic cleavage activities from Caldanaerobius polysaccharolyticus.* J Bacteriol, 2010. **192**(16): p. 4111-21.

16. Oyama, T., et al., *Mutational and structural analyses of Caldanaerobius polysaccharolyticus Man5B reveal novel active site residues for family 5 glycoside hydrolases.* PLoS One, 2013. **8**(11): p. e80448.

17. Chekan, J.R., et al., *Structural and biochemical basis for mannan utilization by Caldanaerobius polysaccharolyticus strain ATCC BAA-17.* J Biol Chem, 2014. **289**(50): p. 34965-77.

18. de Souza, W.R., *Microbial Degradation of Lignocellulosic Biomass*, in *Sustainable Degradation of Lignocellulosic Biomass - Techniques, Applications and Commercialization*, A. Chandel, Editor. 2013, INTECH. p. 40.

19. Aspinall, G.O., *Chemistry of Cell Wall Polysaccharides*, in *The Biochemistry of plants: a comprehensive treatise*. 1980, Academic Press, New York: In: Press J. p. 473-500.

20. Montgomery, R., F. Smith, and H.C. Srivastava, *Structure of Corn Hull Hemicellulose. I. Partial Hydrolysis and Identification of 2-O-(α-D-Glucopyranosyluronic Acid)-D-xylopyranose1,2.* Journal of the American Chemical Society, 1956. **78**(12): p. 2837-2839.

21. Eda, S., A. Ohnishi, and K. KatŌ, *Xylan Isolated from the Stalk ofNicotiana tabacum.* Agricultural and Biological Chemistry, 1976. **40**(2): p. 359-364.

22. Wilkie, K.C.B. and S.-L. Woo, *A heteroxylan and hemicellulosic materials from bamboo leaves, and a reconsideration of the general nature of commonly occurring xylans and other hemicelluloses.* Carbohydrate Research, 1977. **57**: p. 145-162.

23. Yamabhai, M., et al., *Mannan biotechnology: from biofuels to health.* Crit Rev Biotechnol, 2016. **36**(1): p. 32-42.

24. Petkowicz, C.L.D., et al., *Linear mannan in the endosperm of Schizolobium amazonicum.* Carbohydrate Polymers, 2001. **44**(2): p. 107-112.

25. van Zyl, W.H., et al., *Fungal β-mannanases: Mannan hydrolysis, heterologous production and biotechnological applications.* Process Biochemistry, 2010. **45**(8): p. 1203-1213.

26. Dhawan, S. and J. Kaur, *Microbial mannanases: an overview of production and applications.* Crit Rev Biotechnol, 2007. **27**(4): p. 197-216.

27. Cervero, J.M., et al., *Enzymatic hydrolysis and fermentation of palm kernel press cake for production of bioethanol.* Enzyme and Microbial Technology, 2010. **46**(3-4): p. 177-184.

28. Dea, I.C.M. and A. Morrison, *Chemistry and Interactions of Seed Galactomannans.* Adv. Carbohydr. Chem. Biochem., 1975. **31**: p. 241-312.

29. Dey, P.M., *Biochemistry of Plant Galactomannans.* Adv. Carbohydr. Chem. Biochem., 1978. **35**: p. 341-376.

30. Klyosov, A.A., et al., *Structural features of beta-(1-->4)-D-galactomannans of plant origin as a probe for beta-(1-->4)-mannanase polymeric substrate specificity.* Carbohydr Res, 2012. **352**: p. 65-9.

31. Willfor, S., et al., *Spruce-derived mannans - A potential raw material for hydrocolloids and novel advanced natural materials.* Carbohydrate Polymers, 2008. **72**(2): p. 197-210.

32. Aro, N., T. Pakula, and M. Penttilä, *Transcriptional regulation of plant cell wall degradation by filamentous fungi.* FEMS Microbiology Reviews, 2005. **29**(4): p. 719-739.

33. De Vries, J.A., et al., *Enzymic degradation of apple pectins.* Carbohydrate Polymers, 1982. **2**(1): p. 25-33.

34. Pérez, S., K. Mazeau, and C. Hervé du Penhoat, *The three-dimensional structures of the pectic polysaccharides.* Plant Physiology and Biochemistry, 2000. **38**(1-2): p. 37-55.

35. Vincken, J.P., et al., *If homogalacturonan were a side chain of rhamnogalacturonan I. Implications for cell wall architecture.* Plant Physiology, 2003. **132**(4): p. 1781-1789.

36. de Vries, R.P., *Regulation of Aspergillus genes encoding plant cell wall polysaccharide-degrading enzymes; relevance for industrial production.* Appl Microbiol Biotechnol, 2003. **61**(1): p. 10-20.

37. Rahikainen, J.L., et al., *Inhibitory effect of lignin during cellulose bioconversion: the effect of lignin chemistry on non-productive enzyme adsorption.* Bioresour Technol, 2013. **133**: p. 270-8.

38. Warren, R.A., *Microbial hydrolysis of polysaccharides.* Annu Rev Microbiol, 1996. **50**: p. 183-212.

39. Henrissat, B., T.T. Teeri, and R.A. Warren, *A scheme for designating enzymes that hydrolyse the polysaccharides in the cell walls of plants.* FEBS Lett, 1998. **425**(2): p. 352-4.

40. Teeri, T.T., *Crystalline cellulose degradation: new insight into the function of cellobiohydrolases.* Trends in Biotechnology, 1997. **15**(5): p. 160-167.

41. Malgas, S., J.S. van Dyk, and B.I. Pletschke, *A review of the enzymatic hydrolysis of mannans and synergistic interactions between beta-mannanase, beta-mannosidase and alpha-galactosidase.* World J Microbiol Biotechnol, 2015. **31**(8): p. 1167-75.

42. Soccol, C.R., et al., *Bioethanol from lignocelluloses: Status and perspectives in Brazil.* Bioresour Technol, 2010. **101**(13): p. 4820-5.

43. Dias, M.O., et al., *Second generation ethanol in Brazil: can it compete with electricity production?* Bioresour Technol, 2011. **102**(19): p. 8964-71.

44. Lennartsson, P.R., P. Erlandsson, and M.J. Taherzadeh, *Integration of the first and second generation bioethanol processes and the importance of by-products.* Bioresour Technol, 2014. **165**: p. 3-8.

45. Kalscheuer, R., T. Stolting, and A. Steinbuchel, *Microdiesel: Escherichia coli engineered for fuel production.* Microbiology, 2006. **152**(Pt 9): p. 2529-36.

46. Henstra, A.M., et al., *Microbiology of synthesis gas fermentation for biofuel production.* Curr Opin Biotechnol, 2007. **18**(3): p. 200-6.

47. Atsumi, S., T. Hanai, and J.C. Liao, *Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels.* Nature, 2008. **451**(7174): p. 86-9.

48. Alper, H. and G. Stephanopoulos, *Engineering for biofuels: exploiting innate microbial capacity or importing biosynthetic potential?* Nat Rev Microbiol, 2009. **7**(10): p. 715-23.

49. Withers, S.T., et al., *Identification of isopentenol biosynthetic genes from Bacillus subtilis by a screening method based on isoprenoid precursor toxicity.* Appl Environ Microbiol, 2007. **73**(19): p. 6277-83.

50. FitzPatrick, M., et al., *A biorefinery processing perspective: treatment of lignocellulosic materials for the production of value-added products.* Bioresour Technol, 2010. **101**(23): p. 8915-22.

51. Dodds, D.R. and R.A. Gross, *Chemistry. Chemicals from biomass.* Science, 2007. **318**(5854): p. 1250-1.

52. McCoy, M., *Biofuels Center Grows in West.* Chemical & Engineering News, 2007. **85**(12): p. 12.

53. Kabel, M.A., et al., *Effect of pretreatment severity on xylan solubility and enzymatic breakdown of the remaining cellulose from wheat straw.* Bioresour Technol, 2007. **98**(10): p. 2034-42.

54. Rosgaard, L., et al., *Evaluation of minimal Trichoderma reesei cellulase mixtures on differently pretreated Barley straw substrates.* Biotechnol Prog, 2007. **23**(6): p. 1270-6.

55. Rosgaard, L., S. Pedersen, and A.S. Meyer, *Comparison of different pretreatment strategies for enzymatic hydrolysis of wheat and barley straw.* Appl Biochem Biotechnol, 2007. **143**(3): p. 284-96.

56. Somerville, C., *Biofuels.* Curr Biol, 2007. **17**(4): p. R115-9.

57. Ensinas, A.V., J.H.S. Arnao, and S.A. Nebra, *Increasing Energetic Efficiency in Sugar, Ethanol, and Electricity Producing Plants*, in *Sugarcane bioethanol — R&D for Productivity and Sustainability*. 2014: São Paulo, SP, Brazil. p. 583-600.

58. Cardona, C.A., J.A. Quintero, and I.C. Paz, *Production of bioethanol from sugarcane bagasse: Status and perspectives.* Bioresour Technol, 2010. **101**(13): p. 4754-66.

59. Rossell, C.E.V., Zainaghi, G., *Produção de Etanol Combustível pela Hidrólise Enzimática*, in *Álcool Combustível - Série Indústria em Perspectiva*, A.Q.M. Neto, Editor. 2008, IEL/NC: Brasília, DF, Brazil. p. 123-138.

60. Nissen, A.M., et al., *Xylanases for the Pulp and Paper-Industry.* Xylans and Xylanases, 1992. **7**: p. 325-337.

61. Ludwig, W., K.-H. Schleifer, and W.B. Whitman, *Revised road map to the phylum Firmicutes*, in *Bergey's Manual® of Systematic Bacteriology: Volume Three The Firmicutes*, P. Vos, et al., Editors. 2009, Springer New York: New York, NY. p. 1-13.

62. Lee, Y.J., et al., *Description of Caldanaerobius fijiensis gen. nov., sp. nov., an inulin-degrading, ethanol-producing, thermophilic bacterium from a Fijian hot spring sediment, and reclassification of Thermoanaerobacterium polysaccharolyticum and Thermoanaerobacterium zeae as Caldanaerobius polysaccharolyticus comb. nov. and Caldanaerobius zeae comb. nov.* Int J Syst Evol Microbiol, 2008. **58**(Pt 3): p. 666-70.

63. Han, Y., et al., *Biochemical and structural insights into xylan utilization by the thermophilic bacterium Caldanaerobius polysaccharolyticus.* J Biol Chem, 2012. **287**(42): p. 34946-60.

64. Davies, G.J., K.S. Wilson, and B. Henrissat, *Nomenclature for sugar-binding subsites in glycosyl hydrolases.* Biochem J, 1997. **321 (Pt 2)**: p. 557-9.

65. Ye, Y., et al., *Structural insights into the difference in substrate recognition of two mannoside phosphorylases from two GH130 subfamilies.* FEBS Lett, 2016. **590**(6): p. 828-37.

66. Cuskin, F., et al., *The GH130 Family of Mannoside Phosphorylases Contains Glycoside Hydrolases That Target β-1,2-Mannosidic Linkages in Candida Mannan.* Journal of Biological Chemistry, 2015. **290**(41): p. 25023-25033.

67. Nihira, T., et al., *An inverting beta-1,2-mannosidase belonging to glycoside hydrolase family 130 from Dyadobacter fermentans.* FEBS Lett, 2015. **589**(23): p. 3604-10.

68. Tsuda, T., et al., *Characterization and crystal structure determination of beta-1,2-mannobiose phosphorylase from Listeria innocua.* FEBS Lett, 2015. **589**(24 Pt B): p. 3816-21.

69. Zhang, M., et al., *Xylan utilization in human gut commensal bacteria is orchestrated by unique modular organization of polysaccharide-degrading enzymes.* Proc Natl Acad Sci U S A, 2014. **111**(35): p. E3708-17.

70. R Developing Core Team, *R: A language and environment for statistical computing*. 2016, R Foundation for Statistical Computing: Vienna, Austria.

71. Williams, G.J., *Data Mining with Rattle and R: The Art of Excavating Data for Knowledge Discovery*. Use R! 2011: Springer.

72. Eswar, N., et al., *Protein Structure Modeling with MODELLER*, in *Structural Proteomics: High-Throughput Methods*, B. Kobe, M. Guss, and T. Huber, Editors. 2008, Humana Press: Totowa, NJ. p. 145-159.

73. Webb, B. and A. Sali, *Comparative Protein Structure Modeling Using MODELLER.* Curr Protoc Bioinformatics, 2014. **47**: p. 5 6 1-32.

74. Ladeveze, S., et al., *Structural bases for N-glycan processing by mannoside phosphorylase.* Acta Crystallogr D Biol Crystallogr, 2015. **71**(Pt 6): p. 1335-46.

75. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* J Mol Graph, 1996. **14**(1): p. 33-8, 27-8.

76. Gottlieb, H.E., V. Kotlyar, and A. Nudelman, *NMR chemical shifts of common laboratory solvents as trace impurities.* Journal of Organic Chemistry, 1997. **62**(21): p. 7512-7515.

APPENDIX A

Table A.1 – Buffers utilized in protein purification

Buffer	Ingredient concentration (mM)				Glycerol	pH
	HEPES	NaCl	TCEP[1]	Imidazole		
Lysis	100	500	0.5	0	10%	8.0
Binding	100	500	0.5	0	10%	8.0
Wash	50	500	0.5	10	10%	7.5
Elution	50	500	0.5	300	10%	7.5
Storage	50	300	0.5	0	10%	7.5

[1]TCEP: Tris-(2-chloroethyl) phosphate

APPENDIX B

Script containing the R codes utilized for the hierarchical clustering utilizing Ward's method:

```
#Installation of required package for Ward's hierarchical clustering method
install.packages("rattle")
library("rattle")
#Import distance matrix generated by MUSCLE
dm <- read.delim("muscle_matrix.txt", header=TRUE, row.names=1)
#convert it to format read by hclust function
d <- as.dist(dm)
#hierarchical clustering with Ward's method
ward <- hclust(d, "ward.D")
#to export output from hierarchical clustering
##Install and load the Cluster and Tree Conversion (ctc) package from Bioconductor
source("https://bioconductor.org/biocLite.R")
biocLite("ctc")
library("ctc")
##Generate DENDROSCOPE file containing phylogenetic tree
write.table(hc2Newick(ward), file="ward_nw.newark", row.names=TRUE, col.names=TRUE)
#Manually remove the quotes in the generated file or utilize the na.strings function when generating the file
```

APPENDIX C

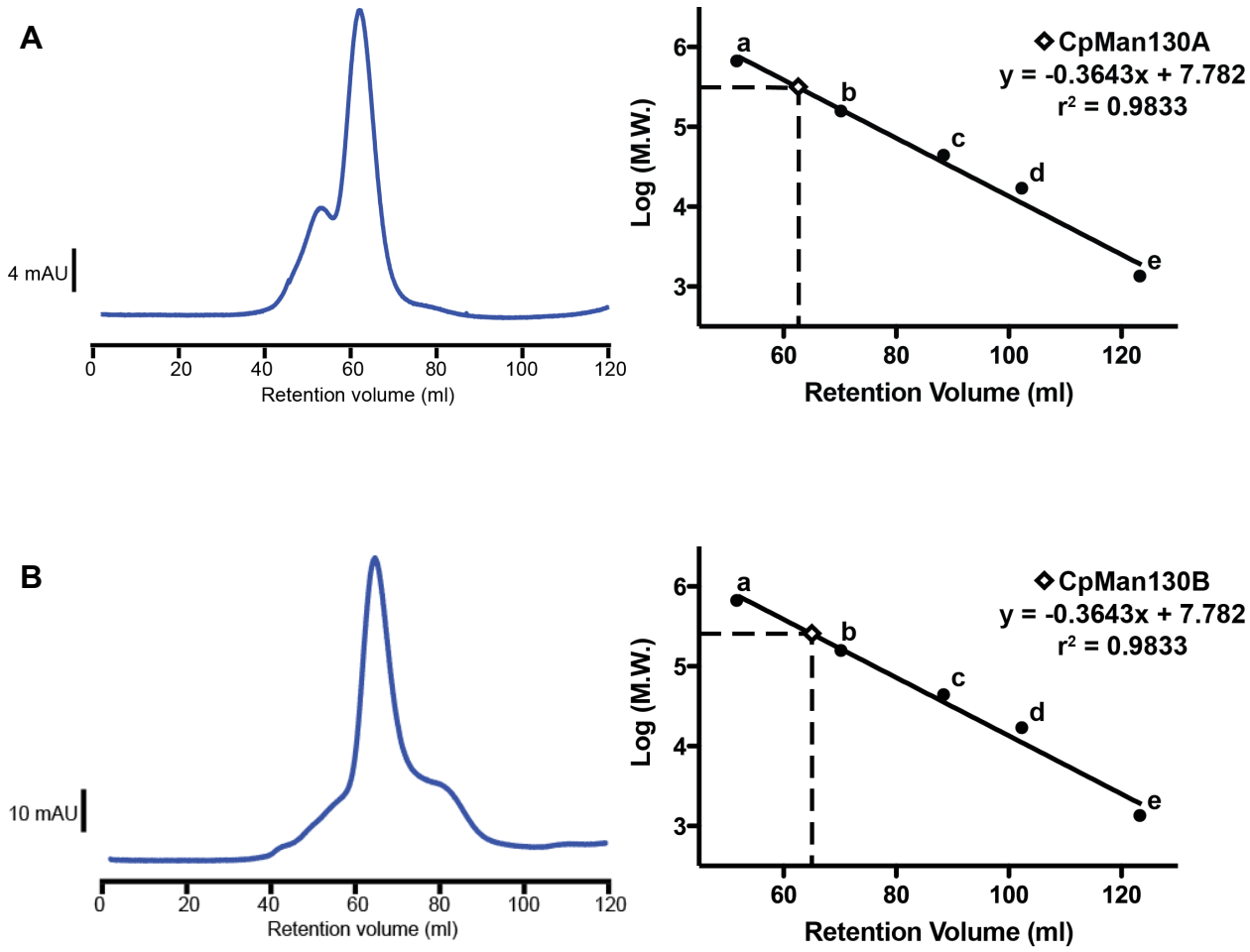


Fig. A.1 – Native molecular weight of the five GH130 proteins
Chromatograms obtained from analytical gel filtration and standard curves containing the standards and marked retention volume for CpMan130A (A), CpMan130B (B), CpMan130C (C), CpMan130D (D) and CpMan130E (E).

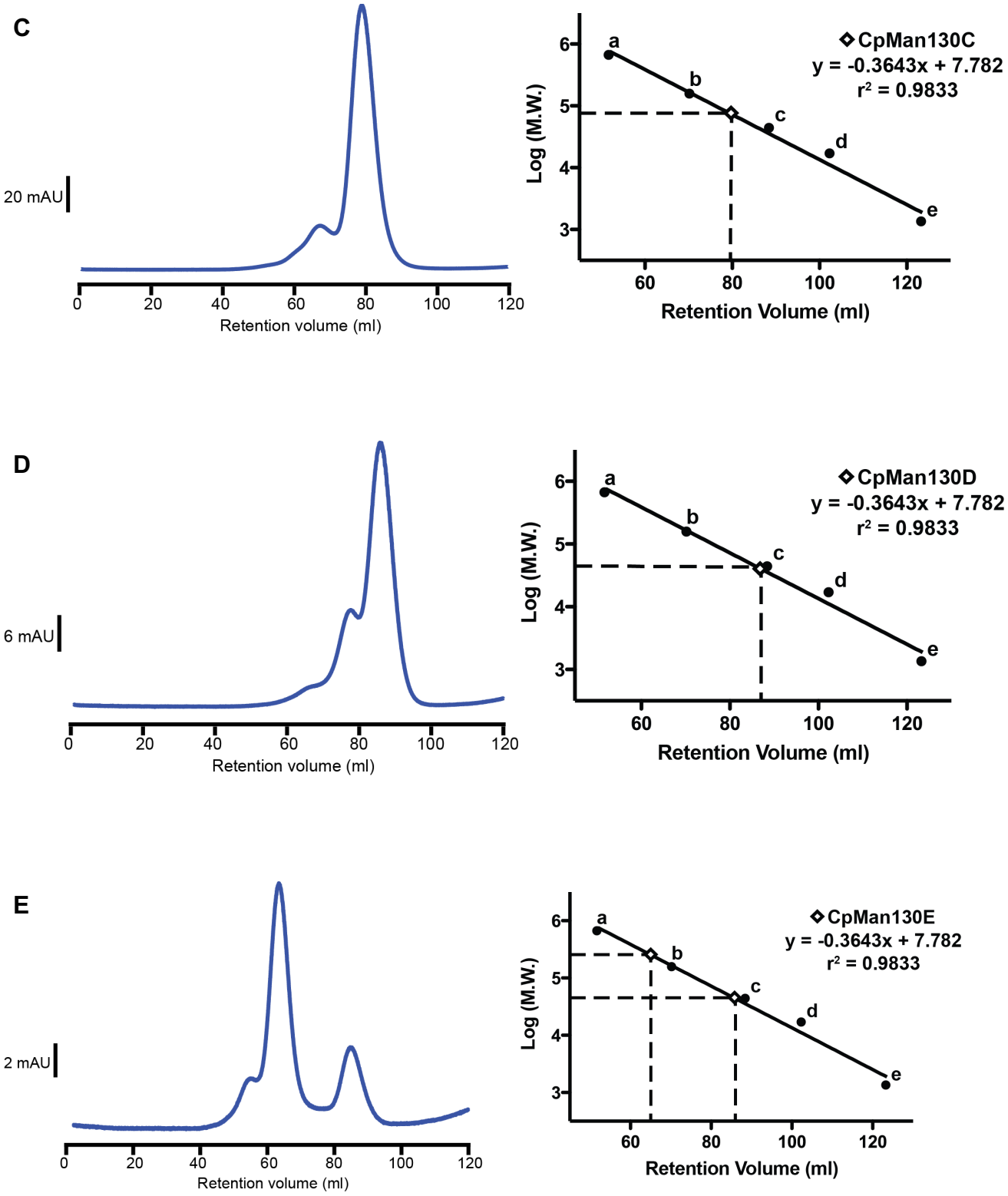


Fig. A.1 – continued.

APPENDIX D

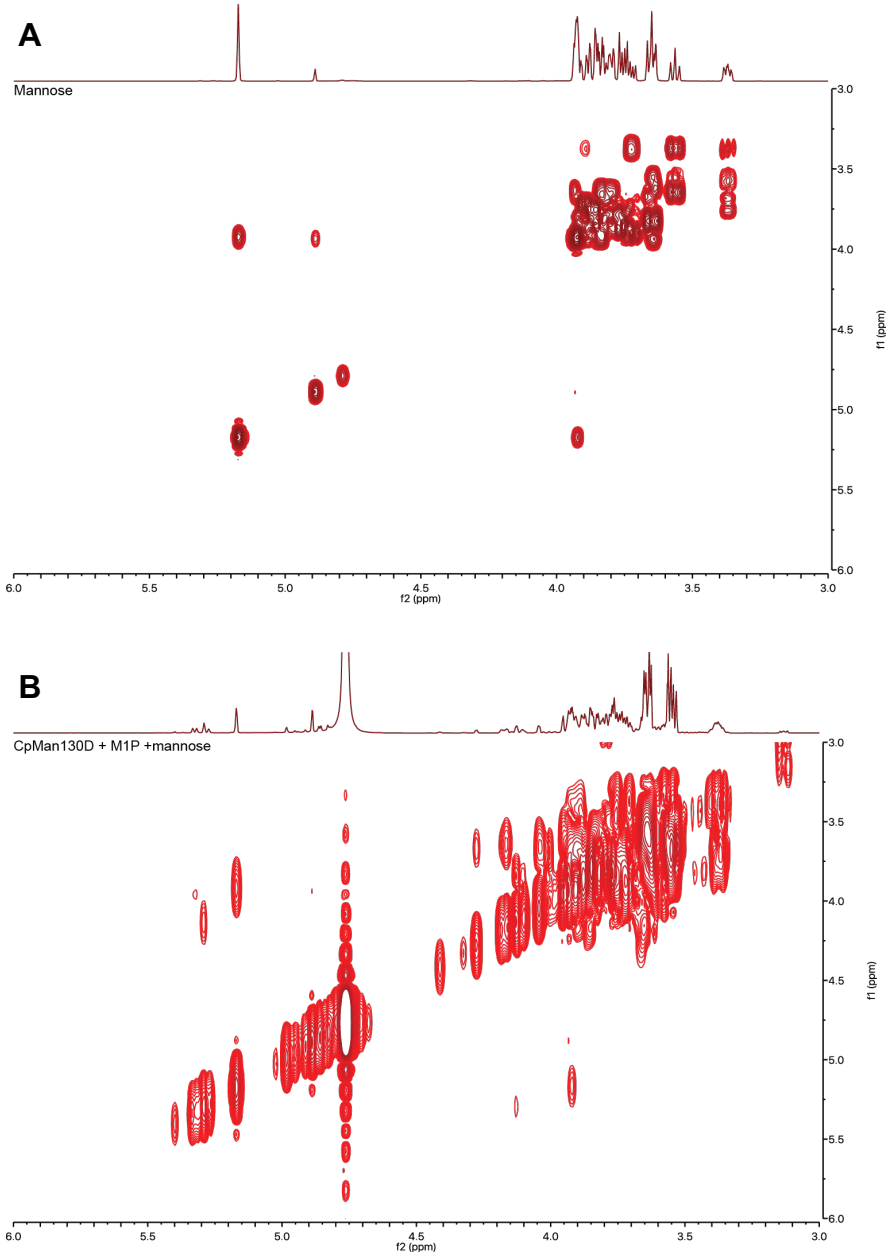


Fig. A.2 – NMR spectra of the products from the reverse reaction of CpMan130D
(A) COSY spectra of mannose in deuterium oxide. (B) COSY spectra of the products from the reverse reaction of CpMan130D after incubation with mannose and mannose-1-phosphate in deuterium oxide. (C) HSQC spectra of mannose in deuterium oxide. (D) HSQC spectra of mannobiose in deuterium oxide. (E) HSQC spectra of the products from the reverse reaction of CpMan130D after incubation with mannose and mannose-1-phosphate in deuterium oxide, showing a proposed structure for the formed β-1,2-linked disaccharide in the inset.

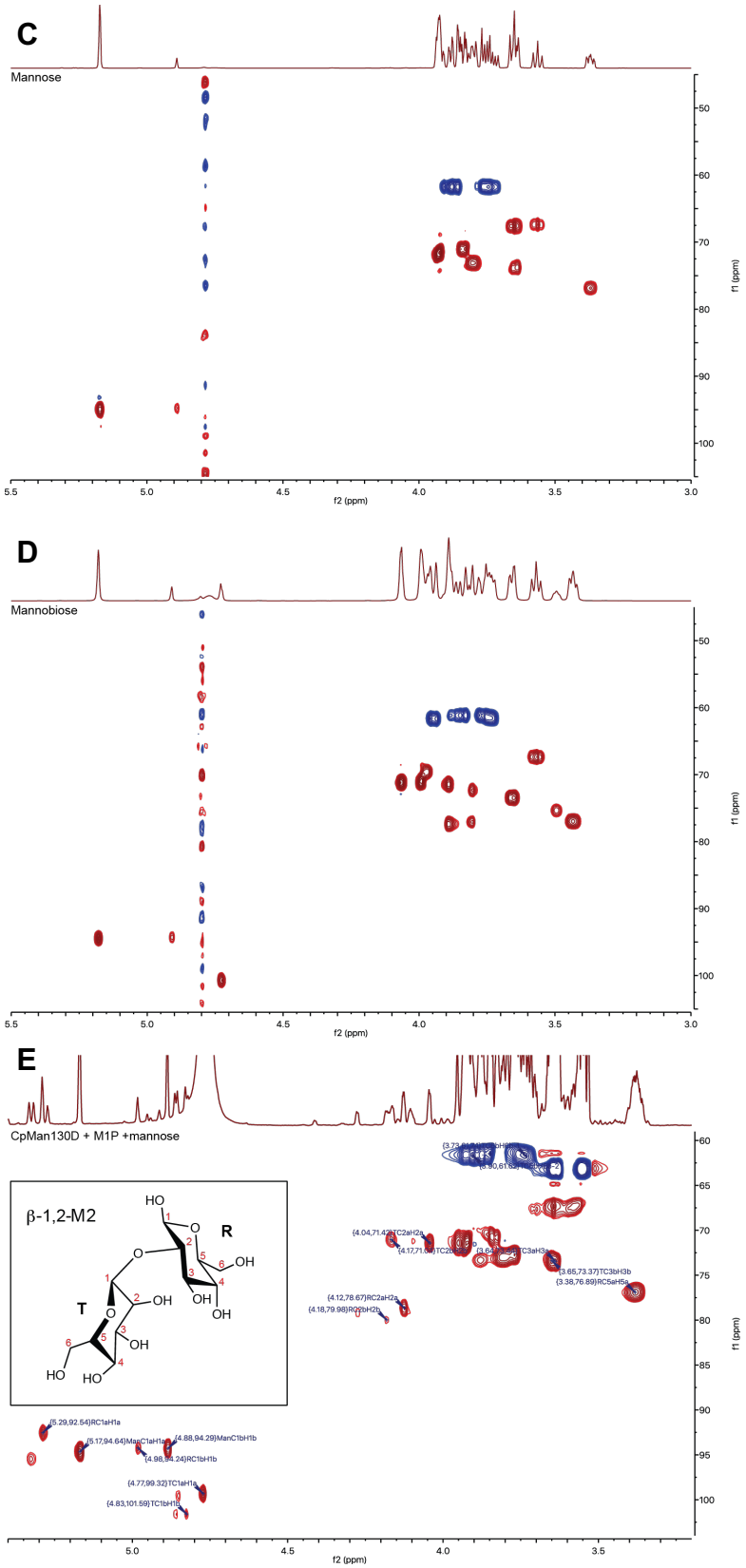


Fig. A.2 – Continued.

APPENDIX E

Table A.2 – Proteins included in subgroup GH130_3

Accession number	Organism	Predicted function
ADG06391.1	*Kyrpidia tusciae* DSM 2912	glycosidase PH1107-related protein
AEE95612.1	*Mahella australiensis* 50-1 B0N	glycosidase related protein
ABP66381.1	*Caldicellulosiruptor saccharolyticus* DSM 8903	glycosidase, PH1107-related
ADL41804.1	*Caldicellulosiruptor obsidiansis* OB47	glycosidase related protein
ADQ03977.1	*Caldicellulosiruptor owensensis* OL	glycosidase related protein
ADQ07826.1	*Caldicellulosiruptor hydrothermalis* 108	glycosidase related protein
ADQ40091.1	*Caldicellulosiruptor kristjanssonii* I77R1B	glycosidase related protein
ACM59662.1	*Caldicellulosiruptor bescii DSM 6725*	glycosidase PH1107-related protein
ADQ46940.1	*Caldicellulosiruptor kronotskyensis* 2002	glycosidase related protein
ADQ07975.1	*Caldicellulosiruptor hydrothermalis* 108	glycosidase related protein
ABP65906.1	*Caldicellulosiruptor saccharolyticus* DSM 8903	glycosidase, PH1107-related protein
BAD40300.1	*Symbiobacterium thermophilum* IAM 14863	glycosylase
CALPO_RS0109450	*Caldanaerobius polysaccharolyticus* ATCC BAA-17	**characterized – CpMan130C**
AEF16970.1	*Thermoanaerobacterium xylanolyticum* LX-11	glycosidase related protein
AFK87322.1	*Thermoanaerobacterium saccharolyticum* JW/SL-YS485	glycosidase related protein
ADL69413.1	*Thermoanaerobacterium thermosaccharolyticum* DSM 571	glycosidase-related
AGB19545.1	*Thermoanaerobacterium thermosaccharolyticum* M0795	putative glycosylase
AAM23370.1	*Caldanaerobacter subterraneus* subsp. *tengcongensis* MB4	predicted glycosylase
ADD01383.1	*Thermoanaerobacter italicus* Ab9	glycosidase PH1107-related protein

Table A.2 – Continued.

Accession number	Organism	Predicted function
ADH59898.1	*Thermoanaerobacter mathranii* subsp. *mathranii* str. A3	glycosidase related protein
ADN53735.1	*Thermoanaerobacter sp.* X514	glycosidase related protein
ABY95802.1	*Thermoanaerobacter pseudethanolicus* ATCC 33223	glycosidase, PH1107-related
ABY93073.1	*Thermoanaerobacter sp.* X514	**characterized – Teth514_1789**
ADV80731.1	*Thermoanaerobacter brockii* subsp. *finnii* Ako-1	glycosidase related protein
AIS51279.1	*Thermoanaerobacter kivui*	β-1,4-mannooligosaccharide phosphorylase
AEM77586.1	*Thermoanaerobacter wiegelii* Rt8.B1	glycosidase related protein
BAD40299.1	*Symbiobacterium thermophilum* IAM 14863	glycosylase
ABP65905.1	*Caldicellulosiruptor saccharolyticus* DSM 8903	glycosidase, PH1107-related protein
ADQ07976.1	*Caldicellulosiruptor hydrothermalis* 108	glycosidase related protein
CALPO_RS0109445	*Caldanaerobius polysaccharolyticus* ATCC BAA-17	**characterized – CpMan130D**
ADL69414.1	*Thermoanaerobacterium thermosaccharolyticum* DSM 571	glycosidase-related
AGB19546.1	*Thermoanaerobacterium thermosaccharolyticum* M0795	putative glycosylase
AEF16969.1	*Thermoanaerobacterium xylanolyticum* LX-11	glycosidase related protein
AFK87323.1	*Thermoanaerobacterium saccharolyticum* JW/SL-YS485	glycosidase related protein
AAM23369.1	*Caldanaerobacter subterraneus* subsp. *tengcongensis* MB4	predicted glycosylase
ADD01382.1	*Thermoanaerobacter italicus* Ab9	glycosidase PH1107-related protein
ADH59897.1	*Thermoanaerobacter mathranii* subsp. *mathranii* str. A3	glycosidase related protein
AEM77585.1	*Thermoanaerobacter wiegelii* Rt8.B1	glycosidase related protein

Table A.2 – Continued.

Accession number	Organism	Predicted function
AIS51278.1	*Thermoanaerobacter kivui*	β-1,4-mannooligosaccharide phosphorylase
ABY95803.1	*Thermoanaerobacter pseudethanolicus* ATCC 33223	glycosidase, PH1107-related
ADN53734.1	*Thermoanaerobacter sp.* X513	glycosidase related protein
ADV80732.1	*Thermoanaerobacter brockii* subsp. *finnii* Ako-1	glycosidase related protein
ABY93074.1	*Thermoanaerobacter sp.* X514	**characterized – Teth514_1788**